

## SEARCH REQUEST FORM

FOR OFFICIAL USE ONLY

Requestor's  
Name:

EUG

Serial

Number:

08/012,269

Date:

9/13/93

Phone:

308-3990

Art Unit:

1813

## Search Topic:

Please write a detailed statement of search topic. Describe specifically as possible the subject matter to be searched. Define any terms that may have a special meaning. Give examples or relevant citations, authors keywords, etc., if known. For sequences, please attach a copy of the sequence. You may include a copy of the broadest and/or most relevant claim(s).

Please search DNA & protein databases  
for Fig. 2A & 2B - They are  
supposed to encode the same gene  
but I don't know where they line  
up -

1. PIR
2. Swiss Prot
3. EMBL
4. GenBank

The effective filing date is 11/7/88 so there  
is no ~~Seq ID~~ <sup>CRF</sup> listing.

RECEIVED  
STIC/STIC  
SEP 13 11 30 AM '93  
U.S. PAT. & TM. OFF.

## STAFF USE ONLY

09-114

Date completed: 09-14-93  
Searcher: Beverly E4994  
Terminal time: 75  
Elapsed time: \_\_\_\_\_  
CPU time: \_\_\_\_\_  
Total time: 85  
Number of Searches: \_\_\_\_\_  
Number of Databases: 1

## Search Site

\_\_\_\_ STIC  
\_\_\_\_ CM-1  
\_\_\_\_ Pre-S

## Type of Search

\_\_\_\_ N.A. Sequence  
\_\_\_\_ A.A. Sequence  
\_\_\_\_ Structure  
\_\_\_\_ Bibliographic

## Vendors

\_\_\_\_ IG Suite  
\_\_\_\_ STN  
\_\_\_\_ Dialog  
\_\_\_\_ APS  
\_\_\_\_ Geninfo  
\_\_\_\_ SDC  
\_\_\_\_ DARC/Questel  
\_\_\_\_ Other

> 0 <  
0| 0 IntelliGenetics  
> 0 <

FastDB - Fast Pairwise Comparison of Sequences  
Release 5.4

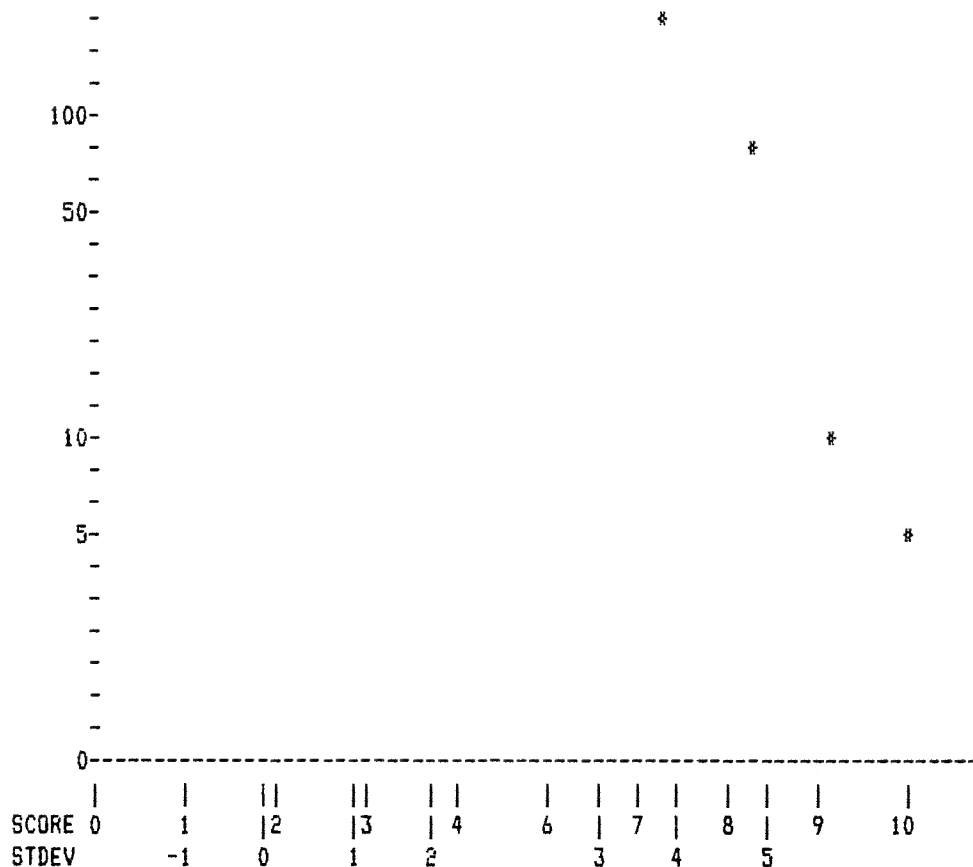
Results file ellis-012-fig2ab-ags.res made by shears on Tue 14 Sep 93 14:54:10-PDT.

Query sequence being compared: ELLIS-012-FIG2AB.PEP (1-256)  
Number of sequences searched: 30847  
Number of scores above cutoff: 4007

Results of the initial comparison of ELLIS-012-FIG2AB.PEP (1-256) with:  
Data bank : A-GeneSeq 11, all entries

100000-  
-  
N -  
U50000-  
M -  
B -  
E -  
R -  
-  
D -  
F10000-  
-  
S -  
E 5000-  
Q -  
U -  
E -  
N -  
C -  
E -  
S 1000-  
\*  
-  
500-

# Best Available Copy



## PARAMETERS

Similarity matrix	Unitary	K-tuple	2
Mismatch penalty	5	Joining penalty	30
Gap penalty	1.00	Window size	32
Gap size penalty	0.26		
Cutoff score	0		
Randomization group	0		

Initial scores to save	40	Alignments to save	15
Optimized scores to save	0	Display context	50

## SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	3	4	1.25

Times:	CPU	Total Elapsed
	00:01:03.08	00:02:09.00

Number of residues:	4048030
Number of sequences searched:	30847
Number of scores above cutoff:	4007

Cut-off raised to 2.  
 Cut-off raised to 3.  
 Cut-off raised to 4.  
 Cut-off raised to 5.  
 Cut-off raised to 6.

The scores below are sorted by initial score.  
 Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Init. Opt.		Sig.	Frame
			Score	Score		
**** 5 standard deviations above mean ****						
1. R04747	Amino acid sequence of modifi	231	10	23	5.60	0
2. R04751	Amino acid sequence of maize	235	10	22	5.60	0
3. R04749	Amino acid sequence of maize	235	10	22	5.60	0
4. R04748	Amino acid sequence of maize	235	10	23	5.60	0
**** 4 standard deviations above mean ****						
5. R28289	HI-30 N-terminal sequence.	20	9	9	4.80	0
6. P91700	Protein increasing pulmonary	23	9	9	4.80	0
7. P91701	Protein increasing pulmonary	35	9	9	4.80	0
8. R30953	Rabbit whey acidic protein.	127	9	16	4.80	0
9. P81110	Sequence of new fusion protei	352	9	34	4.80	0
10. R31046	Rat D1B dopamine receptor.	475	9	18	4.80	0
11. R21082	Dopamine D1 receptor encoded	477	9	19	4.80	0
12. R22546	Truncated Dopamine D1 recepto	479	9	20	4.80	0
**** 3 standard deviations above mean ****						
13. R31224	Transmembrane region of HIV-1	28	8	9	4.00	0
14. R27470	HIV-1 (IIIB) env transmembran	28	8	9	4.00	0
15. R15248	Carbohydrate binding domain #	32	8	9	4.00	0
16. R22089	Human MK protein.	143	8	15	4.00	0
17. P80745	Sequence of AAs 600-750 of HI	150	8	13	4.00	0
18. R24301	Glycopeptide resistance prote	161	8	15	4.00	0
19. P20007	Hybrid human leukocyte interf	187	8	15	4.00	0
20. P20103	Sequence encoded by leukocyte	188	8	15	4.00	0
21. R20564	O-glycosylated IFN-alpha2c.	188	8	15	4.00	0
22. R20549	Human IFNalpha 2C from pAD19B	188	8	15	4.00	0
23. R11802	Sporamin A encoded by the cDN	219	8	17	4.00	0
24. R11356	Alkaline phosphatase-IFN alph	219	8	17	4.00	0
25. P95375	Sequence of lipase of Bacillu	247	8	30	4.00	0
26. P70831	Sequence of lipase of Bacillu	247	8	31	4.00	0
27. R06495	Beta 3 adrenergic receptor.	402	8	36	4.00	0
28. R12395	Transcription activator.	406	8	16	4.00	0
29. R05539	Rat D2 dopamine receptor.	415	8	14	4.00	0
30. R30886	ETb receptor.	442	8	29	4.00	0
31. R10544	D2 dopamine receptor long iso	444	8	15	4.00	0
32. R22499	[GARSYQ]-[Plasminogen 347-541	467	8	35	4.00	0
33. R22032	Truncated human urinary throm	475	8	35	4.00	0
34. R22503	[GARSYQ]-[Plasminogen 347-541	476	8	35	4.00	0
35. R22013	Truncated human thrombomoduli	480	8	35	4.00	0
36. R13877	Thrombin-binding substances (	486	8	35	4.00	0
37. R24400	Recombinant thrombin-binding	494	8	35	4.00	0
38. R10617	Soluble thrombomodulin deriv.	515	8	35	4.00	0
39. R22018	Human thrombomodulin (1-516)	516	8	35	4.00	0
40. R22017	Human thrombomodulin (1-516)	516	8	35	4.00	0

1. ELLIS-012-FIG2AB.PEP (1-256)

R04747 Amino acid sequence of modified 19 kD maize zein e

ID R04747 standard; protein; 231 AA.  
AC R04747;  
DT 05-AUG-1990 (first entry)  
DE Amino acid sequence of modified 19 kD maize zein encoded by clone cZ19A2  
KW Maize zein; lysine substitution.  
OS Maize.  
PN US4885357-A.  
PD 05-DEC-1989.  
PF 21-APR-1988; 184348.  
PR 21-APR-1988; US-184348.  
PA (HERR) Lubrizol Corp. (PUBS)



CC a modified 19 kD zein. It has better nutritional balance than unmodified  
 CC zein (which lacks Lys), but retains the other properties zein - ability  
 CC to form protein bodies within the rough endoplasmic reticulum of the  
 CC host cell, and solubility in alcohol.

SQ Sequence 235 AA;

SQ 37 A; 3 R; 9 N; 1 D; 0 B; 3 C; 39 Q; 1 E; 0 Z; 5 G; 3 H;

SQ 11 I; 48 L; 1 K; 3 M; 14 F; 21 P; 17 S; 5 T; 0 W; 8 Y; 6 V;

Initial Score = 10 Optimized Score = 22 Significance = 5.60  
 Residue Identity = 21% Matches = 32 Mismatches = 98  
 Gaps = 19 Conservative Substitutions = 0

```

60      70      80      90     100     110     120     130
NCNICRVCAGYFRFKKFCSSSTHNAECECIEGFHCLGPQCTRCEKDCRPGQELTKQGCKTCSLGTENDQNGTG
                                     | |  | |
                                     MAAKIFCLIMLLG-LSASAATA
                                     X      10      20

140      150      160      170      180      190      200
VCRPWTNCSLDGRSVL--KTGTTEKDVVCGPPVVSFSPSTTISVTPEGGPGGHSLOVLTFLALTSALLLAL
|  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
SIFP--QCSQAPIASLLPPYLSPAMSSVCENP--ILLPYRIQQAIAAG-----ILPLSPLFLQSSALLQQL
      30      40      50      60      70      80

210      220      230      240      250      X
IFITLLFSVLKWKIRKFKPHIFKQPFKKTGAAQEEEDACSCRCPEEEEGGGGYEL
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
PLVHLL---AQNIR---AQQLQQLVLANLAAYSQQQQFLPFNLALNSAAYLQQQLLPFSQLAAAYPRQ
      90      100      110      120      130 X      140

FLPFNLALNSHAYVQQQLLPFSQLAAVSFA
150      160      170      180

```

### 3. ELLIS-012-FIG2AB.PEP (1-256)

R04749 Amino acid sequence of maize zein encoded by clone

ID R04749 standard; protein; 235 AA.

AC R04749;

DT 05-AUG-1990 (first entry)

DE Amino acid sequence of maize zein encoded by clone cZ19AB1

KW Maize zein; lysine substitution.

OS Maize.

PN US4885357-A.

PD 05-DEC-1989.

PF 21-APR-1988; 184348.

PR 21-APR-1988; US-184348.

PA (LUBR) Lubrizol Corp (PURD).

PI Larkins B, Cuellar RE, Wallace JC;

DR WPI; 90-050879/07.

PT New modified zein contg. lysine residues -

PT with better nutritional balance, prepd. by expressing mutated  
 PT zein gene

PS Disclosure; Fig 4; 18pp; English.

CC The patent concerns a modified 19 or 22 kD zein which includes Lys in the  
 CC internal repeated region of the zein. This is the amino acid sequence of  
 CC a modified 19 kD zein. It has better nutritional balance than unmodified  
 CC zein (which lacks Lys), but retains the other properties zein - ability  
 CC to form protein bodies within the rough endoplasmic reticulum of the  
 CC host cell, and solubility in alcohol.

SQ Sequence 235 AA;

SQ 37 A; 3 R; 9 N; 1 D; 0 B; 3 C; 39 Q; 1 E; 0 Z; 5 G; 3 H;

SQ 11 I; 48 L; 1 K; 3 M; 14 F; 21 P; 17 S; 5 T; 0 W; 8 Y; 6 V;

Initial Score = 10 Optimized Score = 22 Significance = 5.60  
 Residue Identity = 21% Matches = 32 Mismatches = 98

Gaps = 19 Conservative Substitutions = 0

```

60      70      80      90      100     110     120     130
NCNICRVACAGYFRFKKFCSSSTHNAEECEIEGFHCLGPQCTRCEKDCRPGQELTKQGCKTCSLGTENDQNGTG
                                | |  ||  |
                                MAAK1FCL1MLLG-LSASAATA
                                X      10      20

140      150      160      170      180      190      200
VCRPWNTNCSLDGRSVL--KTGTTEKDVVCGPPVVSFSPSTT1SVTPEGGPGGHSLSVLTLFLALTSALLLAL
|  ||  |  ||  |  |  |  |  |  ||  |||  |||  |
SIFP--QCSQAPIASLLPPYLSAMSSVCENP--ILLPYR1QQAIAAG-----ILPLSPLFLQSSALLQQL
      30      40      50      60      70      80

210      220      230      240      250      X
1FITLLFSVLKW1RKKFPH1FKQPFKKTGAAQEEEDACSCRCPEEEEGGGGYEL
||  ||  |  ||  |  |
PLVHLL---AQN1R---AQQLQQLVLANLAAYSQ0000FLPFNQLAALNSAAYLQ0000LLPFSQLAAAYPRO
  90      100      110      120      130 X      140

FLPFNQLAALNSHAYVQ0000LLPFSQLAAVSPA
150      160      170      180

```

#### 4. ELLIS-012-F1G2AB.PEP (1-256)

R04748 Amino acid sequence of maize zein encoded by clone

ID R04748 standard; protein; 235 AA.  
AC R04748;  
DT 05-AUG-1990 (first entry)  
DE Amino acid sequence of maize zein encoded by clone cZ19B1  
KW Maize zein; lysine substitution.  
OS Maize.  
PN US4885357-A.  
PD 05-DEC-1989.  
PF 21-APR-1988; 184348.  
PR 21-APR-1988; US-184348.  
PA (LUBR) Lubrizol Corp (PURD).  
PI Larkins B, Cuellar RE, Wallace JC;  
DR WPI; 90-050879/07.  
DR N-PSDB; 003296.  
PT New modified zein contg. lysine residues -  
PT with better nutritional balance, prepd. by expressing mutated  
PT zein gene  
PS Disclosure; Fig 4; 18pp; English.  
CC The patent concerns a modified 19 or 22 kD zein which includes Lys in the  
CC internal repeated region of the zein. This is the amino acid sequence of  
CC a modified 19 kD zein. It has better nutritional balance than unmodified  
CC zein (which lacks Lys), but retains the other properties zein - ability  
CC to form protein bodies within the rough endoplasmic reticulum of the  
CC host cell, and solubility in alcohol.  
SQ Sequence 235 AA;  
SQ 27 A; 2 R; 10 N; 0 D; 0 B; 3 C; 42 Q; 1 E; 0 Z; 8 G; 2 H;  
SQ 10 I; 45 L; 1 K; 2 M; 15 F; 23 P; 19 S; 9 T; 0 W; 8 Y; 8 V;

Initial Score = 10 Optimized Score = 23 Significance = 5.60  
Residue Identity = 22% Matches = 33 Mismatches = 97  
Gaps = 19 Conservative Substitutions = 0

```

60      70      80      90      100     110     120     130
NCNICRVACAGYFRFKKFCSSSTHNAEECEIEGFHCLGPQCTRCEKDCRPGQELTKQGCKTCSLGTENDQNGTG
                                | |  ||  |
                                MAAK1FCFLMLLG-LSASAATA
                                X      10      20

140      150      160      170      180      190      200

```





200 210 220 230 240 250 X  
 LALTSALLLALIFITLLFSVLKWKIRKKFPHIFKQPFKKTGAAQEEEDACSCRCPQEEEGGGGGYEL  
 |||||  
 AVLPEEEGGGGGLVTKKED  
 X 10 X 20

# 6. ELLIS-012-FIG2AB.PEP (1-256)

P91700 Protein increasing pulmonary surfactant activity.

ID P91700 standard; protein; 23 AA.  
 AC P91700;  
 DT 13-JUN-1990 (first entry)  
 DE Protein increasing pulmonary surfactant activity.  
 KW Pulmonary surfactant; respiratory disorders;  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT Misc-difference 14  
 FT /label=ile, gly or val pref. ile  
 FT Misc-difference 16  
 FT /label=ile, gly or val, pref. gly  
 PN W08900167-A.  
 PD 12-JAN-1989.  
 PF 29-JUN-1988; 00361.  
 PR 01-JUL-1987; SE-027249.  
 PR 22-SEP-1987; SE-036612.  
 PA (KABI) Kabigen Ab.  
 PI Curstedt T, Robertsson B, Jornvall H;  
 DR WPI; 89-039631/05.  
 PT Proteins with pulmonary surfactant activity -  
 PT obtd. from pig lung and human broncho-alveolar lavage or  
 PT amniotic fluid, for treating respiratory disorders.  
 PS Claim 1; Page 16; 24pp; English.  
 CC Proteins, derived from bronchoalveolar lavage and amniotic fluid, can be  
 CC extracted and shown to have pulmonary surfactant activity. Useful in  
 CC treating respiratory disorders, reducing surface tension at air-liquid  
 CC interface.  
 SQ Sequence 23 AA;  
 SQ 1 A; 0 R; 0 N; 0 D; 0 B; 0 C; 0 Q; 0 E; 0 Z; 2 G; 0 H;  
 SQ 1 I; 7 L; 0 K; 1 M; 0 F; 0 P; 0 S; 0 T; 0 W; 0 Y; 9 V;  
 SQ 2 Others;

Initial Score = 9 Optimized Score = 9 Significance = 4.80  
 Residue Identity = 39% Matches = 9 Mismatches = 14  
 Gaps = 0 Conservative Substitutions = 0

X 10 20 30 40 50 60 70  
 MGNNCYNVVVVIVLLLVGCEKVGAVQNSCDNCQPGTFCRKYNPVCKSCPPSTFSSIGGQPNCNICRVCAGYFR  
 ||| ||| ||  
 LLVVVVVVLLVVVXIXGALLMGL  
 X 10 20 X

FKKFC

# 7. ELLIS-012-FIG2AB.PEP (1-256)

P91701 Protein increasing pulmonary surfactant activity.

ID P91701 standard; protein; 35 AA.  
 AC P91701;  
 DT 13-JUN-1990 (first entry)  
 DE Protein increasing pulmonary surfactant activity.  
 KW Pulmonary surfactant; respiratory disorders;  
 OS Sus scrofa.  
 FH Key Location/Qualifiers

FT Misc-difference 1  
 FT /label=leu or phe  
 FT Misc-difference 9  
 FT /label=asn or his  
 FT Misc-difference 26  
 FT /label=ile, gly or val, pref. gly  
 FT Misc-difference 28  
 FT /label=ile, gly or val, pref. gly  
 PN W08900167-A.  
 PD 12-JAN-1989.  
 PF 29-JUN-1988; 00361.  
 PR 01-JUL-1987; SE-027249.  
 PR 22-SEP-1987; SE-036612.  
 PA (KABI) Kabigen Ab.  
 PI Curstedt T, Robertsson B, Jornvall H;  
 DR WPI; 89-039631/05.  
 PT Proteins with pulmonary surfactant activity -  
 PT obtd. from pig lung and human broncho-alveolar lavage or  
 PT amniotic fluid, for treating respiratory disorders.  
 PS Claim 2; Page 16; 24pp; English.  
 CC Proteins, derived from pig lung, can be extracted and shown to have  
 CC pulmonary surfactant activity. Useful in treating respiratory disorders,  
 CC reducing surface tension at the air-liquid interface.  
 SQ Sequence 35 AA;  
 SQ 1 A; 2 R; 0 N; 0 D; 0 B; 2 C; 0 Q; 0 E; 0 Z; 2 G; 0 H;  
 SQ 2 I; 8 L; 1 K; 1 M; 0 F; 2 P; 0 S; 0 T; 0 W; 0 Y; 10 V;  
 SQ 4 Others;

Initial Score = 9 Optimized Score = 9 Significance = 4.80  
 Residue Identity = 32% Matches = 9 Mismatches = 19  
 Gaps = 0 Conservative Substitutions = 0

	X	10	20	30	40	50	60
	MGNNCYNVVVIVLLLVGCEKVGAVQNSCDNCQPGTFCRKYNPVCKSCPPSTFSSIGGQPNICR						
	XRIPCCPVXLKRLLVVVVVLLVVVXIXGALLMGL						
	10	20	30	X			
70	VCAGYFRFKKFC						

#### 8. ELLIS-012-FIG2AB.PEP (1-256)

R30953 Rabbit whey acidic protein.

ID R30953 standard; Protein; 127 AA.  
 AC R30953;  
 DT 07-MAY-1993 (first entry)  
 DE Rabbit whey acidic protein.  
 KW WAP; promoter; heterologous protein production.  
 OS Oryctolagus cuniculus.  
 PN W09222644-A.  
 PD 23-DEC-1992.  
 PF 12-JUN-1992; F00533.  
 PR 12-JUN-1991; FR-007179.  
 PA (INRG) INRA INST NAT RECH AGRONOMIQUE.  
 PI Devinoy E, Houdebine L, Thepot D;  
 DR WPI; 93-018131/02.  
 DR N-PSDB; 034591.  
 PT Heterologous protein prodn. in milk of transgenic mammal - contg.  
 PT structural gene under control of promoter of rabbit acidic whey  
 PT protein, e.g. for human growth hormone  
 PS Disclosure; Fig 5; 38pp; French.  
 CC The expression control elements from at least a 3kb fragment from  
 CC the 3'-end of the complete rabbit WAP gene are fused to a sequence  
 CC encoding a heterologous protein, such as human growth hormone.

erythropoietin, granulocyte colony stimulating factor,  
alpha-antitrypsin, hirudin, urokinase and Factor VIII. The rabbit  
WAP promoter is far more efficient at expressing such proteins in  
primary mammalian epithelial cells (induced by prolactin and  
glucocorticoids) than rat or mouse WAP promoters. The preferred  
regulatory region is a 6.3kb HindIII-BamHI fragment or a 17kb  
HindIII-EcoRI fragment from the region immediately upstream of the  
rabbit WAP gene (The sequence of only the first 1821 bases upstream  
of the first exon is given in the specification).

Sequence 127 AA:

13 A; 6 R; 2 N; 5 D; 0 B; 14 C; 4 Q; 9 E; 0 Z; 6 G; 0 H;  
6 I; 16 L; 4 K; 4 M; 2 F; 12 P; 12 S; 5 T; 1 W; 1 Y; 5 V;

Initial Score = 9 Optimized Score = 16 Significance = 4.80  
Residue Identity = 22% Matches = 18 Mismatches = 58  
Gaps = 3 Conservative Substitutions = 0

140 150 160 170 180 190 200  
GVCPRWNTNCSLDGRSVLKTGTTEKDVVCGPPVVSFSPSTTISVTPEGGPGGHS LQVLT L-FLALTSALLLAL  
| | | || | || |  
MRCLISLALGLL ALEAALALAP  
X 10 20

210 220 230 240 250 X  
IFI--TLLFSVLKWIRKKFP HIFKQPFKKTGA AQQEEDACSCRC PQQEEGGGGGYEL  
|| | || | |  
KFIAPVQVMCPESPSSSEETLCLSDNDCLGSTVCCPSAAGGSCRTPIIVPTPKAGRC PWVQAPMLSQ LCEELS  
30 40 50 60 70 80 90

DCANDIECRGDKKCCFSRCAMRYLEPILESTPQ  
100 110 120

#### 9. ELLIS-012-FIG2AB.PEP (1-256)

P81110 Sequence of new fusion protein contg. alpha-1-micr

ID P81110 standard; protein; 352 AA.  
AC P81110;  
DT 06-DEC-1990 (first entry)  
DE Sequence of new fusion protein contg. alpha-1-microglobulin (AMG)  
DE and the HI-30 region of inter-alpha-trypsin inhibitor (III) light chain  
KW Serine protease; enzyme; pancreatitis; atherosclerosis;  
KW chronic inflammation; therapy; elastase.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT Protein 20..202  
FT /label=AMG  
FT Protein 206..350  
FT /label=HI-30  
FT Domain 226..282  
FT /label=I  
FT Domain 283..352  
FT /label=II  
FT Misc-difference 291..292  
FT /note="Differs from the protein sequence of HI-30  
FT purified from urine"  
FT Misc-difference 343  
FT /note="Differs from the protein sequence of HI-30  
FT purified from urine"  
PN EP-255011-A.  
PD 03-FEB-1988.  
PF 20-JUL-1987; 110461.  
PR 29-JUL-1986; US-891469.  
PA (MILE) Miles Laboratories Inc.  
PI Kauney JF, Kotick MP, Polazzi JO;  
DE WP1: 88-270242/05

DR N-PSDB; n81432.  
PT New DNA sequence coding for fusion protein contg. alpha-microglobulin -  
PT and inter-alpha-trypsin inhibitor, useful for treating excessive  
PT elastase prodn.  
PS Disclosure; p; English.  
CC A fusion protein of the ITI light chain comprising AMG and HI-30 is  
CC claimed. ITI is serine protease, potentially used for treating excessive  
CC release of hydrolytic enzymes, esp. elastase, in conditions such as  
CC pancreatitis, atherosclerosis and chronic inflammation.  
SQ Sequence 352 AA;  
SQ 21 A; 18 R; 13 N; 12 D; 0 B; 16 C; 13 Q; 28 E; 0 Z; 36 G; 4 H;  
SQ 15 I; 27 L; 18 K; 10 M; 14 F; 19 P; 20 S; 26 T; 5 W; 15 Y; 22 V;

Initial Score = 9 Optimized Score = 34 Significance = 4.80  
Residue Identity = 19% Matches = 46 Mismatches = 168  
Gaps = 23 Conservative Substitutions = 0

```

      10      20      X 30      40      50      60      70
MGNNCYNVVVIVLLLVGCEKVGAVQNSCDNCQPGTFCRKYNPVCKSCPPSTFSSIGGQPNCNICR--VCAGY
                                |   |   |   |   |
                                MRSLGALLLLLSACLAVSAGVPVPTPPDNIGVQENFNISRIYGWYN
                                X   10   20   30   40

      80      90      100     110     120     130     140
FRFKKFCSSSTHNAECECIEGFHCLGPQCTRCCKDCRPGQELTKQGCKTCSLGTENDQNGTGVCRPWNTCSLD
    |           ||   |   |           |   |   |           |
LAIGSTCPWLKKIMDRMTVSTLVLGEGATEAE-ISMTSTRWRKGVCEETS-GAYEKTDTDG-----KFLY
    50      60      70      80      90     100

      150     160     170     180     190     200     210
GRSVLKTGTTEKDVVCGPPVVSFSPSTTISVTEGGPGGHSLOVLTFLALTSALLLALIFITLLFSVLKWI
    |   |   ||   |           |           |   |   ||   |   |
HKS-KWNITNESYVVHTTYDEYAIFLTKKFSRHHGPTITAKLYGRAPQLRET--LL--QDFRVVAQGV--GI
110      120     130     140     150     160     170

      220     230     240     250     X
RKKFPHFIFKQPFKKTGAAGQEDACSCR-----CPQEEEGGGGGYEL
                |   |   |           ||||| |||
PEDSIFTMADRGECPVGEQEPEPILIPRVPRVLPQEEEGSGGGQLVTEVTKKEDSCQLGYSAGPCMGMTSR
    180     190     200     210     220     230     240

YFYNGTSMACETFQYGGCMGNGNNF
    250     260     270

```

10. ELLIS-012-FIG2AB.PEP (1-256)  
R31046 Rat D1B dopamine receptor.

ID R31046 standard; Protein; 475 AA.  
AC R31046;  
DT 26-MAY-1993 (first entry)  
DE Rat D1B dopamine receptor.  
KW PCR; amplify; degenerate; primer; TM; transmembrane region; human; D1;  
KW dopamine; receptor; probe; rat; pBLUESCRIPT II SK+; testis; DR5; D1B;  
KW genomic library; lambdaDASH II; Kozak; consensus sequence; V-15.  
OS Rattus rattus.  
PN W09218533-A.  
PD 29-OCT-1992.  
PF 16-APR-1992; U03187.  
PR 16-APR-1991; US-686591.  
PA (UYDU-) UNIV DUKE.  
PI Caron MG; Jarvie KR; Tiberi M;  
DR WPI; 93-036060/04.  
DR N-PSDB; 035148.  
PT Cloned gene encoding rat D1b dopamine receptor - used to screen  
PT cnds. for receptor activity or in receptor binding assays.

PS Disclosure; Page 25-28; 39pp; English.

CC This sequence represents rat D1B dopamine receptor. The DNA sequence encoding this polypeptide was isolated using the primer sequences given in Q35146-47. These oligomers are degenerate primers corresponding to the 5th and 6th transmembrane (TM) regions of the human D1 dopamine receptor. These primers were used to amplify sheared human DNA and the amplification products were subcloned into the sequencing vector pBLUESCRIPT II SK+. A 230bp fragment (V-15) was found to correspond to the 5th TM region, the 3rd intracellular loop and the 6th TM region. V-15 was used as a template for the synthesis of a 32P-labeled probe. This probe was used to screen a rat testis genomic library in lambdaDASH II. One isolated clone (DR5) had an open reading frame of 1425 bp (475 amino acids) which contained the full coding sequence for rat D1B-dopamine receptor. The predicted encoded protein has a molecular weight of 52834. The putative initiator methionine was selected on the basis of the best Kozak consensus sequence found in frame with the remainder of the coding block and preceded by a stop codon.

SQ Sequence 475 AA;

SQ 41 A; 26 R; 17 N; 18 D; 0 B; 15 C; 16 Q; 27 E; 0 Z; 27 G; 7 H;

SQ 34 I; 42 L; 13 K; 12 M; 25 F; 26 P; 40 S; 28 T; 13 W; 11 Y; 37 V;

Initial Score = 9 Optimized Score = 18 Significance = 4.80  
 Residue Identity = 23% Matches = 28 Mismatches = 72  
 Gaps = 17 Conservative Substitutions = 0

```

100      110      120      130      140      X 150      160
PQCTRCEKDCRPGQELTKQGCKTCSLGTFNQNGTGVCRPWTNCSLDGRSVLKTG--TTEKDVVCG-PPVVS
                                | | | |
                                MLPPGRNRTAQPARLGLQRQLA
                                X      10      20

170      180      190      200      210      220      230
FSPSTTISVTPGGPGGHSLOQLTLFLALTSALLLALIFITLLFSVL--KWIRKKFPHIFK-QPFKKTGAA
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QVDAPAGSATPLG-----PAQVVTAGL-LT--LL---IVWTLLGNVLVCAIVRSRHLRAKMTNIFIVSLAV
30          40          50          60          70          80

240      250      X
QEEDACSCRCPEEEGGGGGYEL
| |
SDFVALLVMPWKAVAEVAGVWPFGTCDIHWAFDIMCSTASILNLCIISVDRYWAISRPFYERKNTQVA
90          100      X 110      120      130      140      150

```

L

# 11. ELLIS-012-FIG2AB.PEP (1-256)

R21082 Dopamine D1 receptor encoded by clone GL-30.

ID R21082 standard; Protein; 477 AA.

AC R21082;

DT 20-MAY-1992 (first entry)

DE Dopamine D1 receptor encoded by clone GL-30.

KW G-protein-coupled receptor; Parkinson's Disease; schizophrenia;

KW tardive dyskinesia; dopamine D1-beta receptor subtype.

OS Homo sapiens.

FH Key Location/Qualifiers

FT Domain 42..66

FT /label= transmembrane

FT /note= "I"

FT Domain 78..101

FT /label= transmembrane

FT /note= "II"

FT Domain 117..174



## 12. ELLIS-012-FIG2AB.PEP (1-256)

R22546 Truncated Dopamine D1 receptor encoded by pseudoge

ID R22546 standard; Protein: 479 AA.  
 AC R22546;  
 DT 20-MAY-1992 (first entry)  
 DE Truncated Dopamine D1 receptor encoded by pseudogene clone GL-39.  
 KW G-protein-coupled receptor; Parkinson's Disease; schizophrenia;  
 KW tardive dyskinesia; dopamine D1-beta receptor subtype.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT Modified\_site 7..9  
 FT /label= glycosylation  
 FT Domain 42..66  
 FT /label= transmembrane  
 FT /note= "I"  
 FT Domain 78..101  
 FT /label= transmembrane  
 FT /note= "II"  
 FT Domain 117..138  
 FT /label= transmembrane  
 FT /note= "III"  
 FT Domain 158..174  
 FT /label= transmembrane  
 FT /note= "IV"  
 FT Misc\_difference 190  
 FT /note= "corresponds to nonsense codon -  
 FT i.e. protein is truncated"  
 FT Domain 226..248  
 FT /label= transmembrane  
 FT /note= "V"  
 FT Domain 296..317  
 FT /label= transmembrane  
 FT /note= "VI"  
 FT Domain 339..362  
 FT /label= transmembrane  
 FT /note= "VII"  
 FT Misc\_difference 457  
 FT /note= "corresponds to nonsense codon"  
 PN W09200986-A.  
 PD 23-JAN-1992.  
 PF 10-JUL-1991; U04858.  
 PR 10-JUL-1990; U5-551448.  
 PA (NEUR-) NEUROGENETIC CORP.  
 PI Weinshank RL, Hartig PR;  
 DR WPI; 92-056815/07.  
 PT Nucleic acid sequences encoding human dopamine D1 receptor - and  
 PT anti-sense oligo-nucleotide(s), useful in treating and diagnosing  
 PT abnormal D1 receptor expression e.g. dementia, etc.  
 PS Disclosure; Fig 2; 90pp; English.  
 CC Clone GL-39 encodes a truncated (and therefore inactive) dopamine  
 CC D1 receptor having strong homology to the full-length receptor  
 CC encoded by GL-30 (see Q21082).  
 SQ Sequence 479 AA;  
 SQ 41 A; 21 R; 22 N; 19 D; 0 B; 17 C; 16 Q; 23 E; 0 Z; 23 G; 8 H;  
 SQ 30 I; 40 L; 11 K; 16 M; 24 F; 33 P; 41 S; 23 T; 14 W; 12 Y; 43 V;  
 SQ 2 Others;

Initial Score = 9 Optimized Score = 20 Significance = 4.80  
 Residue Identity = 25% Matches = 31 Mismatches = 72  
 Gaps = 19 Conservative Substitutions = 0

90 100 110 120 130 140 150 160

GEHCL GPACTGCEKDCRPQDEL TKACCKTCLCTENBANCYUCCPDTNCCI DCPENL KTTCTEKNUUCGCP

KMTGRM

13. ELLIS-012-FIG2AB.PEP (1-256)

R31224 Transmembrane region of HIV-1 (IIIB) env.

```

ID      R31224 standard; peptide; 28 AA.
AC      R31224;
DT      18-MAY-1993 (first entry)
DE      Transmembrane region of HIV-1 (IIIB) env.
KW      Human immunodeficiency virus; fusion protein; transmembrane anchor;
KW      env; T1; T2; TH4.1; epitope.
OS      Synthetic.
PN      W09222641-A.
PD      23-DEC-1992.
PF      12-JUN-1992; U05107.
PR      14-JUN-1991; US-715921.
PR      11-JUN-1992; US-897382.
PA      (VIRO-) VIROGENETICS CORP.
PI      Cox WI, Paoletti E, Tartaglia J;
DR      WPI; 93-018128/02.
PT      Modified recombinant virus with inactivated non-essential genetic
PT      functions - comprises e.g. vaccinia or avipox virus, used as HIV
PT      vaccine
PS      Example 32; Page 102; 159pp; English.
CC      Fusion peptides expressed by recombinant poxviruses include the 51
CC      amino acid N-terminal portion of HIV-1 (IIIB) env, residues 1-50
CC      (plus an initiating Met). The signal sequence is followed by the
CC      T1, T2 and TH4.1 epitopes separated from the signal, each other, and
CC      the anchor sequence where present, by a cleavable linker region up to
CC      5 amino acids in length. The anchor domain is a 28 amino acid trans-
CC      membrane region of HIV-1 (IIIB) env (sequence shown).
CC      See also R31218-26.
SQ      Sequence 28 AA;
SQ      1 A; 3 R; 1 N; 0 D; 0 B; 0 C; 1 Q; 0 E; 0 Z; 4 G; 0 H;
SQ      3 I; 4 L; 0 K; 1 M; 2 F; 0 P; 1 S; 0 T; 0 W; 0 Y; 7 V;

```

```
Initial Score      =      8  Optimized Score =      9  Significance = 4.00
Residue Identity = 32%  Matches           =      9  Mismatches  = 19
Gaps              =      0  Conservative Substitutions =      0
```

X 10 20 30 X 40 50 60 70  
MGNNCYNVVVIVLLLVGCEKVGAVQNSCDNCQPGTFCRKYNPVCKSCPPSTFSSIGGQPNICRVCAGYFR  
II III I II I  
LFIMIVGGLVGLRIVFAVLSSVNRVRQG  
X 10 20 X



14. ELLIS-012-FIG2AB.PEP (1-256)

R27470 HIV-1 (IIIB) env transmembrane region.

ID R27470 standard; Protein; 28 AA.

AC R27470;

DT 24-FEB-1993 (first entry)

DE HIV-1 (IIIB) env transmembrane region.

KW T1; T2; TH4.1; epitope; HIV-1; env; transmembrane anchor domain;

KW vP1060; vP1061; vCP154; vCP148; fusion peptide; signal sequence;

KW cleavable linker; H6 promoter; polymerase chain reaction; PCR;

KW vaccinia virus.

OS Synthetic.

FH Key Location/Qualifiers

FT Binding\_site 62

FT /note= "Transmembrane anchor region binding site"

PN W09215672-A.

PD 17-SEP-1992.

PF 09-MAR-1992; U01906.

PR 07-MAR-1991; US-666056.

PR 11-JUN-1991; US-713967.

PR 06-MAR-1992; US-847951.

PA (VIRO-) VIROGENETICS CORP.

PI Cox WI, De Taisne C, Francis J, Gettig RR, Johnson GP;

PI Linbach KJ, Norton EK, Paoletti E, Perkus ME, Pincus SE;

PI Riviere M, Tartaglia J, Taylor J.

DR WPI; 92-331718/40.

PT Vaccine comprises recombinant, attenuated pox-virus - use for

PT vaccinating against viral infections such as rabies, hepatitis B,

PT HIV, HSV, EBV, CMV, mumps etc.

PS Disclosure; Page 327; 456pp; English.

CC The sequences given in Q35846-51 and R27468-70 were used for the  
 CC expression of two fusion peptides containing the T1, T2 and TH4.1  
 CC epitopes of HIV-1 env with and without a transmembrane anchor domain  
 CC from HIV-1 env. Plasmids vP1060, vP1061, vCP154 and vCP148 were  
 CC generated to express a fusion peptide consisting of the signal  
 CC sequences from HIV-1 env coupled to sequences corresponding to the T1,  
 CC T2 and TH4.1 epitopes of HIV-1 env by cleavable linker. vP1060 and  
 CC vCP154 differ from vP1061 and vCP148 in that the former recombinant  
 CC viruses express the fusion protein along with sequences corresponding  
 CC to the transmembrane region of HIV-1 env. The HIV-1 (IIIB) env signal  
 CC region and vaccinia virus H6 promoter are derived by polymerase chain  
 CC reaction (PCR). The remainder of the coding regions for construction  
 CC without the transmembrane region were also produced by PCR. For the  
 CC version containing the transmembrane region the 3' end of the  
 CC amplification product was altered to accommodate the transmembrane region.  
 CC See also Q35501-864.

SQ Sequence 28 AA;

SQ 1 A; 3 R; 1 N; 0 D; 0 B; 0 C; 1 Q; 0 E; 0 Z; 4 G; 0 H;

SQ 3 I; 4 L; 0 K; 1 M; 2 F; 0 P; 1 S; 0 T; 0 W; 0 Y; 7 V;

Initial Score = 8. Optimized Score = 9 Significance = 4.00  
 Residue Identity = 32% Matches = 9 Mismatches = 19  
 Gaps = 0 Conservative Substitutions = 0

X 10 20 30 X 40 50 60 70  
 MGNNCYNVVVIVLLLVGCEKVGAVQNSCDNCQPGTFCRKYNPVCKSCPPSTFSSIGGQPNICRVCAGYFR  
 || ||| ||  
 LFIMIVGGLVGLRIVFAVLSVYVNRVRQG  
 X 10 20 X

80

FKKFCSTHNAE

15. ELLIS-012-FIG2AB.PEP (1-256)

R15248 Carbohydrate binding domain #5.

ID R15248 standard; Protein: 32 AA.  
AC R15248;  
DT 12-FEB-1992 (first entry)  
DE Carbohydrate binding domain #5.  
KW cellulose; CBD; hemicellulosic substrate;  
KW Trichoderma reesei; cellulase; terminal A region.  
PN W09117244-A.  
PD 14-NOV-1991.  
PF 08-MAY-1991; DK0124.  
PR 09-MAY-1990; DK-001158.  
PA (NOVO ) NOVO NORDISK A/S.  
PI Woldike HF, Hagen F, Hjort CM, Hastrup S.  
DR WPI; 91-353766/48.  
PT New fungal (hemi)cellulose degrading enzymes - for prodn. of liq.  
PT fuel gas and feed protein, have specified carbohydrate binding domain  
PS Claim 20; Page 45; 73pp; English.  
CC This CBD is homologous to a terminal A region of Trichoderma reesei  
CC cellulases and effects binding of a protein to an insoluble  
CC cellulosic or hemicellulosic substrate. It is one of ten specific  
CC CBD's (see R15244-R15253) which correspond to the generic CBD  
CC formulae in R15242 and R15243. The CBD is incorporated into a fusion  
CC protein comprising a catalytic domain from a cellulase, e.g. a  
CC Bacillus endoglucanase, and optionally comprising a linking B domain  
CC from e.g. a fungal endoglucanase.  
SQ Sequence 32 AA;  
SQ 1 A; 1 R; 2 N; 0 D; 0 B; 5 C; 7 Q; 0 E; 0 Z; 6 G; 0 H;  
SQ 0 I; 1 L; 0 K; 0 M; 0 F; 1 P; 2 S; 2 T; 3 W; 1 Y; 0 V;

Initial Score = 8 Optimized Score = 9 Significance = 4.00  
Residue Identity = 29% Matches = 10 Mismatches = 22  
Gaps = 2 Conservative Substitutions = 0

```

      60      70      80      90     100 X   110      120
SSIGGQPNCNICRVCAGYFRFKKFCSSSTHNAECECIEGFHCLGPQCTRCEKDCRPGQELTKQGCKTCSLGT
                                     ||   ||   ||   ||
                                     WGQCGGQ--GWQGP TC CSQGT C
                                      X      10      20

```

```

      130     X 140      150      160      170      180
NDQNGTGVCRPWTNCSLDGRSVLKTGTTEKDVVCGPPVVSFSPSTTISVTPEGGPGGHSLOV
||
RAQNGWYSQCLN
      30 X

```

The scatter plot displays the relationship between the number of users (N) and the number of files (F) for various systems. The y-axis represents the number of users (N) and ranges from 0 to 10000. The x-axis represents the number of files (F) and ranges from 0 to 1000. Data points are marked with asterisks (\*). The plot shows a general trend where the number of files increases as the number of users increases, with some systems having a high number of users but a low number of files.

System	Files (F)	Users (N)
U	~100	~5000
M	~200	~4000
B	~300	~3000
E	~400	~2000
R	~500	~1500
D	~600	~1000
F	~700	~800
S	~800	~600
E	~900	~500
Q	~1000	~400
U	~1100	~300
E	~1200	~200
N	~1300	~150
C	~1400	~100
E	~1500	~80
S	~1600	~60
	~1700	~50
	~1800	~40
	~1900	~30
	~2000	~20
	~2100	~15
	~2200	~10
	~2300	~8
	~2400	~6
	~2500	~5
	~2600	~4
	~2700	~3
	~2800	~2
	~2900	~1
	~3000	~0.5



18. Q10263	pZ130 contg. Calgene lambda 1	4383	119	941	6.01	0
19. Q11415	Ryanodine receptor gene.	15464	119	989	6.01	0
**** 5 standard deviations above mean ****						
20. N91839	Pasteurella multocida toxin g	4380	116	494	5.83	0
21. Q21645	3' coding sequence of P.falci	1297	115	531	5.77	0
22. N71064	Gene encoding Plasmodium viva	1908	115	707	5.77	0
23. N40166	Sequence of A.awamori glucoam	3408	115	976	5.77	0
24. Q10883	30kD TNF inhibitor precursor	2088	114	728	5.70	0
25. Q10955	Encodes human 55kD TNF-bindin	2111	114	730	5.70	0
26. Q06285	Human Tumour Necrosis Factor-	2141	114	730	5.70	0
27. Q12215	Type I TNF receptor.	2176	114	725	5.70	0
28. Q34941	Calgene Lambda 140 genomic cl	4383	113	941	5.64	0
29. Q35143	Calgene lambda 140/pZ130 DNA	4383	113	944	5.64	0
30. Q10319	Calgene lambda 140 genomic cl	4383	113	941	5.64	0
31. Q20532	Sequence of clone lambdaAPCP1	2256	112	919	5.58	0
32. Q10014	Clone lambda APCP168i4 of bet	2256	112	919	5.58	0
33. N80604	Lambda APCP168i4, amino acids	2256	112	917	5.58	0
34. Q05086	Sequence encodes NAP-2 gene a	2949	112	939	5.58	0
35. N91050	Sequence encoding novel amylo	2949	112	937	5.58	0
36. Q24442	Encodes truncated TNF-alpha 5	474	110	204	5.46	0
37. Q24441	Encodes truncated TNF-alpha 5	608	110	203	5.46	0
38. N90907	Glutamine synthesis gene.	1200	110	503	5.46	0
39. Q06282	Plasmid Tumour Necrosis Facto	1334	110	427	5.46	0
40. Q03599	Human liver cytochrome P-450	1818	110	453	5.46	0

# 1. ELLIS-012-FIG2AB.SEQ (1-2350)

Q21695 Plasma membrane proton ATPase.

ID Q21695 standard; DNA; 2933 BP.  
AC Q21695;  
DT 02-JUN-1992 (first entry)  
DE Plasma membrane proton ATPase.  
KW Antifungal agents; H+ ATPase; ss.  
OS Candida albicans.  
FH Key Location/Qualifiers  
FT CDS 151..2842  
FT /\*tag= a  
FT /product= H+ ATPase  
PN EP-472286-A.  
PD 26-FEB-1992.  
PF 18-JUL-1991; 306542.  
PR 18-JUL-1990; US-555123.  
PA (MERI ) MERCK & CO INC.  
PI Kurtz MB, Marrinan JA;  
DR WPI; 92-066496/09.  
DR P-PSDB; R21580.  
PT New gene for evaluating antifungal agents - encodes Candida  
PT albicans plasma membrane H-adenosine:tri:phosphatase  
PS Claim 2; Page 8; 25pp; English.  
CC A large, single colony of Candida albicans ATCC 10261 was cultured  
CC and chromosomal DNA extracted. The DNA was digested with restriction  
CC enzymes and fragments probed with a fragment isolated from plasmid  
CC B1138 contg. the Saccharomyces cerevisiae plasma membrane ATPase  
CC (PMA1) gene in the pUC18 vector. Multiple restriction enzyme digests  
CC showed the C. albicans DNA to be homologous to the S. cerevisiae  
CC fragment. A library of C. albicans genomic DNA was constructed,  
CC (rich in the DNA encoding the plasma membrane proton ATPase) using  
CC strain WD-1 and inserted into pEMBL3-23. A positive clone of 12-  
CC 15 kb was ligated into the YEp24 vector, and transformed in E. coli  
CC K-12 strain DH5 alpha. Recombinant plaques were isolated and  
CC sequenced, showing a gene of 2.7 kb. The gene can be used to  
CC transform non-pathogenic yeast which can be used to evaluate agents  
CC capable of perturbing C. albicans plasma membrane H+ ATPase activity.  
CC The gene also provides a means for producing large amounts of the  
CC plasma membrane protein.

50 Sequence 2933 BP; 758 A; 518 C; 633 G; 1024 T;

Initial Score = 146 Optimized Score = 764 Significance = 7.69  
Residue Identity = 47% Matches = 940 Mismatches = 768  
Gaps = 267 Conservative Substitutions = 0

```

  460      470      480      490      500 X  510      520
GGACTGCAGGCCTGGCCAGGAGCTAACGAAGCAGGGTTGCAAAACCTGTAGCTTGGGAACATTTAATGACCA
                                ||  | |||| | |
                                TCT----ATCATTTGTTAA---
                                X      10

  530      540      550      560      570      580      590
GAACGGTACTGGCGTCTGTGACCCCTGGACGAAGTCTCTAGACGGAAGTCTGTGCTTAAGACCGGGAC
  || |  | | | ||| | | | | | | | | | | | | | | | |
-----TATT----TATTTATACCAAGCACCA-----TATAAATACCTAGTTTTTTTTTTTTTTTG----
      20      30      40      50      60

  600      610      620      630      640      650      660
CACGGAGAAGGACGTGGTGTGTGGACC---CCCTGTGGTGAGCTTCTCTCCAGTACCA-CCATTTCTGTGA
  | | | | | | | | | | | | | | | | | | | | | | | |
TTTGTAAATCACTTTTTTTTCAATCTTTGTTTTGGTTAATTAATCT-TAAGAATAAGGATTTTTATAT
  70      80      90     100     110     120     130

  670      680      690      700      710      720
CTCCAGA---GGGAGGACCA-GGAGGGCAC-TCCTTGCAGG-TCCTTACCTTGT-TCC--TGGCGCTGACA
  | | | | | | | | | | | | | | | | | | | | | | | |
ATATATAAACCATGAGTGCTACTGAACCAACCAACGAAAGGTTGATAAAATCGTCTCCGATGATGAAGACC
140     150     160     170     180     190     200     210

  730      740      750      760      770      780      790
TCGGCTTTG-CTGCT--GGCCCTGATCTTCATTACT--CTCCTGTTCTGTGCTCAAATGGATCAGGA-AA
  | | | | | | | | | | | | | | | | | | | | | | | |
AAGACATTGACCAATTAGTCGCTGATTACAAT-CTAACCAGGTGCT-GGTGATGAAGAAGAAGAGGAGGA
      220     230     240     250     260     270     280

  800      810      820      830      840      850      860
AAATTCCCCCACATATTCAAGCAACCATTTAAGAAG-ACCACTGG-AGCAG-CTCAAGAGGAAGATGCTTGT
  ||| | | | | | | | | | | | | | | | | | | | | |
AAATGACTCTTC--CTTCAA--AGCCCTCCAGAGAATTATTGAAACTGACCCAAG----AGTTGTT-T
      290     300     310     320     330     340

  870      880      890      900      910      920      930
AGCTGCCGATG--TCCACAGGAAGAAGAAGGAGGAGGAGGCTATGAGCTGTGATGTACTATCCTAGGAG
  ||| |||| | | | ||||| | | | | | | | | | | | |
GACTGATGATGAAGTCACCAAAAGAAGAAAGA-GATACGGTTTGAATCAAATG-GCTGAAGAA--CAAGAAA
      350     360     370     380     390     400     410

  940      950      960      970      980      990     1000
ATGTGTGGGCCGAAA-CCGAGAAGCACTAGGACCCACCATCCTGTGGAACAGCACAAGCAACCCACACC
  | | | | | | | | | | | | | | | | | | | | | | | |
ACTTG-GTTCTTAAATTCGTATGTTCTTTG-----TTGGTCCAATTCAATTCGTTATGGAAGCCGCTGC-
      420     430     440     450     460     470

 1010     1020     1030     1040     1050     1060     1070
CTGTCTTACACATCATCCTAGATGA-TGTGTGGGCGCGCACCTCATCCAAG-TCTCTTAAAGCTAACAT
  ||| | | | | | | | | | | | | | | | | | | | | |
-TGTTTTGGCTGCTGCT-TTGAAGATTGGGTCGATTTCCGTGTTATCTGTGCTTTATTGTTATTGAATGCT
      480     490     500     510     520     530     540

 1080     1090     1100     1110     1120     1130     1140
AT--TTGTCTTTA-CCTTTTTTA--AATCTTTTTTAAATTTAAATTTATGTGTGAGTGTGCTGCTGC
  | | | | | | | | | | | | | | | | | | | | | | | |
TTTGTGGTTTTATCCAAGAATACCAAGCTGTTCT-ATTGTCGAT-GAAT-TGAAAAAGACTTTGCCCAAC
      550     560     570     580     590     600     610
```

1150 1160 1170 1180 1190 1200  
CTGTATGCACACGTGTGTGTG---TGTGTGTGTG-TGACTCCTCATGCCCTGAGGAGGTGAGAAGAGAAA  
| | | | | | | | | | | | | | | | | | | | |  
---TCTGCTCTTGT-TGTTAGAAACGGTCAATTAGTTGAAATCCAGCTAAC-GAAGTTGTTCCAGGTGATA  
620 630 640 650 660 670 680

1210 1220 1230 1240 1250 1260  
GGGTTGGTTCATAAGA--ACTGGA--GTTAT-----GGATGGCTG--TGAGCCGGNNGATAGGTCGGG  
| | | | | | | | | | | | | | | | | | | | |  
-TCTTG---CAATTGGAAGACGGTACC GTTATTCCAAC TGATGGTAGAATTGTTTCTG-AAGATTGTTTGT  
690 700 710 720 730 740

1270 1280 1290 1300 1310 1320 1330  
AC--GGAGACCTGTCTTCTTATTTTAAC--GTGACTGTATAATAAAAAAAAAATGATATTTCCGGAATTGTA  
| | | | | | | | | | | | | | | | | | | | |  
ACAAGTTGATCAATCTGC-TATT---ACTGGTGAATCTTTAGCTGTCGACAAAGAAGT---GGTGACTCTT  
750 760 770 780 790 800 810

1340 1350 1360 1370 1380 1390 1400  
GAGATTGTCCTGACACCCTT---CTAGTTAATGATCTAAGAGGAATTGTTGATACGTAGTATACTGTATAT  
| | | | | | | | | | | | | | | | | | | | |  
GTTACTCTTCTTCTACTGTTAAGACTGGTGAAGCCTTTATGA-TTGTACTGCTAC-TGGT-GACTCTACTT  
820 830 840 850 860 870 880

1410 1420 1430 1440 1450 1460  
GTGTATGTATA--TG-T-ATATGTATATATAAG---ACTCTTTTACTGTCAAAGTCAACCTAGAGTGTCTG-  
| | | | | | | | | | | | | | | | | | | | |  
TCGTCGGTAGAGCTGCTGCTTTGGTTAAACAAGCTTCCGCTGGTACTGGTCATTTCA--CTGAAGTCT-TGA  
890 900 910 920 930 940 950

1470 1480 1490 1500 1510 1520 1530  
--GTTACCAGGTCAATTTTTATTGGACATTTTACGTCACACACACACACACACACACACACGTTTATAC  
| | | | | | | | | | | | | | | | | | | | |  
ACGGTATTGGTACTACCTTGTGGTCTTT----GTCATTGTTACTTTGTTGGTGGTTTGGGTTGCTTGTTC  
960 970 980 990 1000 1010 1020

1540 1550 1560 1570 1580 1590 1600  
TAC-GTACTGTTATCGGTATTCTACGT CATATAATGGGATAGGGTAAAGGAAACCAAGAGTGAGTGAT--  
| | | | | | | | | | | | | | | | | | | | |  
TACAGAACCGTTA---GAATTGTTT--CA-ATCTTGAGATACACTTTAGCCATCACTATTATTG-GTGTTC  
1030 1040 1050 1060 1070 1080

1610 1620 1630 1640 1650 1660 1670  
-ATTATTGTG-GAGGTGACAGACTACCCCTT--TGGGTACGTAGGGACAGACCTCCTTCGGACTGTCTAAA  
| | | | | | | | | | | | | | | | | | | | |  
AGTTGGTTTGCAGCTGTC--GTTACCACTACCATGGCT--GTCGGTGCTG-CTTACTTGGCCAAGAAACAA  
1090 1100 1110 1120 1130 1140 1150

1680 1690 1700 1710 1720 1730  
---ACTCCCCTTAGA-AGTCT--CGTCAAGTTCCCGGACGAAGAGGACAGAGGACACAGTCCGAAAAGTT  
| | | | | | | | | | | | | | | | | | | | |  
GCTATTGTCCAAAAATTGTCTGCCATTGAATCTTTGGCTGGTGTGAAATCTTGTGTTCCGATAAAACCGGT  
1160 1170 1180 1190 1200 1210 1220

1740 1750 1760 1770 1780 1790 1800  
ATTTTTCC---GGCAAA---TCCTTTCCCTGTTTCGTGACACTCCACCCCTTGTGGA--CACTTGAGTGTC  
| | | | | | | | | | | | | | | | | | | | |  
ACTTTGACCAAGAACAATTGTCCTTGAC--GAACCAT-ACACTGTTGAAGGTGTTGAACAGATGACT-TG  
1230 1240 1250 1260 1270 1280 1290

1810 1820 1830 1840 1850 1860  
ATCCTTGCGCCGGAAGGTGAGGTGGTACCGTCTGTAG--GGGCGGGGA-GACA--GA-----GCCGCGGGG  
| | | | | | | | | | | | | | | | | | | | |  
AT-GTTG-ACTGCTTGTGTTAGCTGCTTCTAGAAAGAAGAGGGTTTGGATGCCATTGATAAGGTTTCTTGA  
1300 1310 1320 1330 1340 1350 1360

```

1870      1880      1890      1900      1910      1920      1930
GAGCTACGAGAATCGACTCACAGGGCGCCCGGGCTTCGC--AAATGAAACTTTTTTAATCTCACAAGTTTC
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
AATCTTTGATCAACTACCA-AGAGCTAAAGCTGCTTTGCCAAATACAAGTTATTGAATTCCAACCTTTC
1370      1380      1390      1400      1410      1420      1430

1940      1950      1960      1970      1980      1990      2000
G-TCCGGGCTCGGCGGACCTATGGCGTCGATCCTTATTACCTTATCCTGGCGCAAGATAAAACAACCAAAA
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
GATCCTGTCTCCAAGAAAGT-TACTG-CTA---TTGTTGAATCA-CCAG-----AAGGTGAAAGAATTATTT
1440      1450      1460      1470      1480      1490

2010      2020      2030      2040      2050      2060      2070
GCCTTGACTCCGGTAC-TAATTCTCCCTGCCGGCCCCCGTAAGCATAACGCGGCGATCTCCACTTTAAGAAC
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
GTGTTAAGGGTGCCCCATTATTTCGTCTTAAAGACTGTTG-AAG-ATGACCACCCAATC-CCA---GAAGA--
1500      1510      1520      1530      1540      1550

2080      2090      2100      2110      2120      2130      2140
CTGGCCGCG--TTCTGCCTGGTCTCGCTTTTCTGTAACGGTCTTACAAAAGTAATTAGTTC-TTGCTTTCAG
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
-TGTCCACGAAAACTACCAAAACACCCTTGCCGAA---TTTGCTTCCAGAGGT-TTCAGATCTTTGGGTGTTG
1560      1570      1580      1590      1600      1610      1620

2150      2160      2170      2180      2190      2200      2210
CCTCCAAG-CTTCTGCTAGTCTATGGCAGCATCAAGGCTGGTATT-TGCTACGGCTGACCGCTACGCCGCCG
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
CCAGAAAGAGAGGTGAAGGTCACCTGGGA-AATTTTG---GGTATTATGCCATGTATG---GATCCAC-----
1630      1640      1650      1660      1670      1680

2220      2230      2240      2250      2260      2270
CAATA-AGGGTACTGGGCGGC--CCGTC---GAAG---GCCC-TTTGGTTTCAGAAACCAAGG--CCCCC
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
CAAGAGATGATACT--GCTGCCACAGTCAATGAAGCTAGAAGATTAGGTTTAAGAGTTAAGATGTTAACTGG
1690      1700      1710      1720      1730      1740      1750

2280      2290      2300      2310      2320      2330      2340
TCATACC-AACGTTTCGACTTTGATTCTTGCCGGTACGTGGTGGTGGTGCCTTAGCTCTT----TCTCGAT
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
TGATGCCGTTGGTATTGCTAAAGAACTTGTGCTCAATTAGGTTTGGGTAC--TAACATTTACGATGCCGAC
1760      1770      1780      1790      1800      1810      1820

X
AG-TTAGAC
| | | | |
AGATTAGGTTTGTCCGGTGGTGGTGACATGGCTGGTCTGAAATTGCTGATTTCGTTGA
1830      1840      1850      1860      1870      1880

```

## 2. ELLIS-012-FIG2AB.SEQ (1-2350)

023313 DNA encoding masking protein high polymer unit pre

ID 023313 standard; DNA; 5136 BP.

AC 023313;

DT 19-AUG-1992 (first entry)

DE DNA encoding masking protein high polymer unit precursor MPU-P.

KW Transforming growth factor beta; TGF-beta; mammalian cancer; ss.

OS Rattus rattus.

FH Key Location/Qualifiers

FT CDS 1..5136

FT /\*tag= a

FT misc\_feature 2209..4722

FT /\*tag= b

FT /note= "N2514, encodes P838"

FT misc\_feature 61..5136

FT /\*tag= c



FT /note= "N5076, encodes P1692"  
PN J04066597-A.  
PD 02-MAR-1992.  
PF 29-JUN-1990; 173679.  
PR 29-JUN-1990; JP-173679.  
PA (NAKA/) NAKAMURA T.  
DR WPI; 92-120902/15.  
DR P-PSDB; R22461.  
PT Masking protein high polymer unit - combines with transforming  
PT growth factor beta produced by mammalian cancer cells to inhibit  
PT then  
PS Claim 13; Page 9; 25pp; Japanese.  
CC The sequence codes for the precursor (MPU-P) of a masking protein  
CC high polymer unit (MPU). The high polymer subunit MPU binds to  
CC transforming growth factor (TGF) beta produced by mammalian cancer  
CC cells. It may be used to inactivate the cancer cells and thus is  
CC useful in the treatment of human cancers.  
CC See also 023314 and 023315.  
SQ Sequence 5136 BP; 1267 A; 1348 C; 1423 G; 1098 T;

Initial Score = 141 Optimized Score = 942 Significance = 7.38  
Residue Identity = 48% Matches = 1139 Mismatches = 938  
Gaps = 282 Conservative Substitutions = 0

```

                                     X      10      20
                                     ATGTCCATGAACTGCTGA--GT
                                     |||  ||  ||||  |||  ||
CCCGATGTGTGTAGGGACGGCCGCTGCATCAACACTCCTGGGGCCTCCGATGCCGAAT--ACTG-TGACAGT
2870      2880      2890      2900      2910      2920      2930

          30      40      50      60      70      80
GGATA-----AACAGCACGGGATATCTCTGTCTA-AAGGAATATT-ACT-ACACCAGGAAAAGGACACATT
|| ||      ||  ||||  || ||  || || || || || || || || || || || || || || || ||
GGGTACCGGATGTCACGACGGGGCCACTGTGAGGATATCGATGAGTGTCTGACCCCAAGTACCTGTCCCGAG
2940      2950      2960      2970      2980      2990      3000

          90      100      110      120      130      140      150
CGACAA-CAGGAAAGGAGCCTGTACAGAAAACCACAGTGTCTGTGCATGTGACATTTCGCCATG---GGA
||||  |  ||  |||  |  |||  ||||  ||||  ||||  |  ||  ||||  |||
GAACAATGCGTGAATTCCCCAGGTTT---TTACCAGTGTGTGCCCTGCACAGAAGGGTT--CCGTGGCTGGA
3010      3020      3030      3040      3050      3060      3070

          160      170      180      190      200      210
A---ACAAGTGTACAACTGGTGGTCAATTGTGCTGCTAGTGGGCTGTGAGAA-GGTGGGAGC-----C
|  ||||  ||  ||||  |  ||  ||||  ||  ||||  ||||  ||||  ||  |||  ||  ||
ATGGACAA-TGCCTCGATGTGGACG--AGTG-CCTGCAGCCAAAGGTCTGTACCAATGGTTCTGCACCAAC
3080      3090      3100      3110      3120      3130      3140

          220      230      240      250      260      270      280
GTGCAGAACTCC----TGTGATAACTGTGAGCCTGG-TACTTTCTGCAGAAAATACAATCCAGTCTG-CAAG
|| |  ||||  ||||  ||||  || ||  || ||  || ||  || ||  || || || || || || || ||
CTGGAAGGCTCCTACATGTG-TTCCTGCCACAAGGGCTAC-AGCCCCACACCAGACCATAGACACTGTCAAG
3150      3160      3170      3180      3190      3200      3210

          290      300      310      320      330      340
A----GCTGCCCTCCAAGTACCTTCTCC--AGCATAGGTGGACAGCCGAAGTGTAAACATCTGCAGAGTGTGT
|  ||  ||  || ||  || || || || || || || || || || || || || || || || || || || ||
ATATTGATG-AATGTGAGCAAGGGAACCTGTGCATGAACGGGCAGTGCAA---AAACA-CTGACGGCTCCTT
3220      3230      3240      3250      3260      3270

          350      360      370      380      390      400      410
GCAGGCTATTTTCAGGTTCAAGAGTTTTGCTCCTCTACCCACAACGCGGAGTGTGAGTGCATTGAAGGATTTC
|  ||  ||  || ||  || || || || || || || || || || || || || || || || || || || ||
CCGGTGATCTGTGG-GCAGGGCTATCAGCT-GTCAGCGGCTAAAGACCAATGTGAAGATATTGACGAATGC
3280      3290      3300      3310      3320      3330      3340
```

420 430 440 450 460 470 480  
CATTGCTTGGGGCCACAG--TGCACAGATGTG-AAAAGGACTGCAGGC-CTGGCCAGGAGCTAACGAAGCA  
| | | | | | | | | | | | | | | | | | | | | |  
GAGCAC-CGTACCTCTGCTCTACGGGCAGTGCAGGAACACAGAGGGCTCCTTCCAGTGTGTTGTGCAACCA  
3350 3360 3370 3380 3390 3400 3410

490 500 510 520 530 540 550  
GGGTGCAAA--ACCTGTAGCTTGGGAACATTTAATGACCAGAACGGTACTGGCGTCTGTGACCCCTGGACG  
| | | | | | | | | | | | | | | | | | | | | |  
GGGTTACAGAGCATCTGT-GCTTGGAGAC---CACTG-CGAGGATATCAATGAATGCT-TGGA----GGAC-  
3420 3430 3440 3450 3460 3470 3480

560 570 580 590 600 610 620  
AACTGCTCTCTAGACGGAAGTCTGTGC-TTAAGA---CCGGGACCACGGAGAAGGACGTG-GTGTGTGGAC  
| | | | | | | | | | | | | | | | | | | | | |  
-AGTAGTGTCTGCCAGGGAGGTGACTGCATCAATACAGAGGGTCCTATGA-CTGCACGTGCCCGGATGGAC  
3490 3500 3510 3520 3530 3540 3550

630 640 650 660 670 680 690  
CCCCTGTG-GTGAGCTTCTCTCCAGTACCACCATTTCTGTGACTCCAGAGGGAGGACCAGGAGGGCACTCC  
| | | | | | | | | | | | | | | | | | | | | |  
TCCAGCTGAATGA-CAATAAGGGCTGTCAAGACATTAATGAATGTGCACAGCCAGGACTCTGTGCAC-CTCA  
3560 3570 3580 3590 3600 3610 3620

700 710 720 730 740 750 760  
T-TGCAGGTCCTTACCTTGTTCCTGGCGCTGACATCGGCTTTG-CTGCTGGCC--CTGATCTTCATTACTCT  
| | | | | | | | | | | | | | | | | | | | | |  
TGGGGAGTGTCTAAAC--ACACAAGGCTC---ATTCCACTGTGTCTG-TGAACAAGGGTTCTCCAT--CTCT  
3630 3640 3650 3660 3670 3680

770 780 790 800 810 820  
CC---TGTTTC--TCT-GTGCTCAAATGGATCAG-GAAAAAATTCCCCACATATTCAAGCAACCATTTAAGA  
| | | | | | | | | | | | | | | | | | | | | |  
GCAGATGGTCGTACTTGTGAAGATATTGATGAGTGTGTTAAACAACACTGTGTGTGACAGTCACGGCTTCTG-  
3690 3700 3710 3720 3730 3740 3750

830 840 850 860 870 880 890  
AGACCACTGGAGCAGCTCAAGAGGAAGATGCTTGTGTA-----GCTGCCGATGTCCACAGGAAGAAGAAGGAGG  
| | | | | | | | | | | | | | | | | | | | | |  
TGACAACACAGCCGGCTCTTTCCGCTGCCTCTGTTATCAGGGCTTTCAAGCCCCACAGGATGGGCAAGG-GT  
3760 3770 3780 3790 3800 3810 3820

900 910 920 930 940 950 960  
AGGAGGAGGCTATGAGCTGTGATGTACTATCCTAGGAG-ATGTGTGGGCCGAAACCGAGAAGCACTAGGACC  
| | | | | | | | | | | | | | | | | | | | | |  
GTGTGGATGTGAACGAATGTGAAGTGC--TCAGTGGTGTATGTGGGAGGCTTTCTGTGAA-AATGTGGAAG  
3830 3840 3850 3860 3870 3880 3890

970 980 990 1000 1010 1020 1030  
CCACCATCCTGTG-GAACAGCACAAAGCAACCCACCACCCTGTTCTTACACATCATCCTAGATGATGTGTGG  
| | | | | | | | | | | | | | | | | | | | | |  
GGTCCTTCTGTGCGTGTGTGCCGATGAGAACAGGA----GTACAGCCCCATGA--CTGG--GCAGTGTCC  
3900 3910 3920 3930 3940 3950

1040 1050 1060 1070 1080 1090  
---GCGGCGACCT---CATCCAAGT----CTCTTCTAACGCTAA-CATATTTGTCTTTACCTTTTTTAAATC  
| | | | | | | | | | | | | | | | | | | | | |  
CTCCCGGGCTACTGAAGATTGAGTGTGGATCGTC-AGCCCAAAGAAGAAAAGAAGGAGTGTATTATAATC  
3960 3970 3980 3990 4000 4010 4020 4030

1100 1110 1120 1130 1140 1150 1160  
TTTTTTTAAATTTAAATTTTATGTGTGTGAGTGTGTTTGCCTGCC--TGTATGCACACGTGTGTGTGTGTGTG  
| | | | | | | | | | | | | | | | | | | | | |  
TCAAT---GATGCCA---GTCTCTGTGATAACGTGCTGCCCCAACGTACCAACAAGAGTG-CTG-CTG  
4040 4050 4060 4070 4080 4090

1170 1180 1190 1200 1210 1220 1230  
TGTGTGACACTCCTGATGCCTGAGGAGGTCAGAAGAGAAAGGGTTGGTT--CCA-TA-AG--AACTG--GAG  
| | | | | | | | | | | | | | | | | | | | | |  
TACATCGGGCGCC---GGCTGGGGA-GACAATTGTGAGATCTTCCCTTGCCAGTCCAGGGGACTGCTGAG  
4100 4110 4120 4130 4140 4150 4160

1240 1250 1260 1270 1280 1290  
TTAT-GGATGGCTGTGAGCCGNNNGATAGGT-----CGGGACGGAGACCTGTCTTCTTATTTTAACGTGA  
| | | | | | | | | | | | | | | | | | | | | |  
TTCTCGGA--AATGTGCCCTAGAGGAAAAGGTTTGTCCCTGCTGGAGA---ATCCTCTTACGAAACCGGTG  
4170 4180 4190 4200 4210 4220

1300 1310 1320 1330 1340 1350 1360  
CTGTATAATAAAAAAAAAAATGA-TATTTTC--GGGAATTGTAGAG--ATTGTCCTGACACCCTTCTAGTTAAT  
| | | | | | | | | | | | | | | | | | | | | |  
GTGAGAACTACAAAGATGCTGACGAATGCCTGCTGTTGGAGAGGAAATCTGCAAAAAC-----GGTTACT  
4230 4240 4250 4260 4270 4280 4290

1370 1380 1390 1400 1410 1420 1430  
GATCTAAGAGGAATTGTTGATACGTAGTATACTGTATATG--TGTATGTA-TATGTATATGTATATATAAGA  
| | | | | | | | | | | | | | | | | | | | | |  
GTTTGAACACTCAGCCTGGGTATGAATGCTACTGCA-AGGAAGGGACATACTACGATCCTGT-CAAATTACA  
4300 4310 4320 4330 4340 4350 4360

1440 1450 1460 1470 1480 1490  
CTCTTTTACTGTCAAAGTCAACCTAGA--GTGCTGCTGTA-CCAGGTCAATTTTATT-GGACATTTTACGT  
| | | | | | | | | | | | | | | | | | | | | |  
GTGTTTGTATGATGGATGAATGCCAAGACCCTAACAGTTGTATCGATGGCCAGTGTGTTAATACAGAGGGC-T  
4370 4380 4390 4400 4410 4420 4430

1500 1510 1520 1530 1540 1550 1560  
CACACACACACACACACACACACACACGTTTATACTACGTACTGTTATCGGTATTCTAC-GTCATAT-AA  
| | | | | | | | | | | | | | | | | | | | | |  
CTTACAACCTGCTTTTGCACCCACCCAATGGTCCCTGGATGCCCT-CTGAGAAGAGATGTGTGCAGCCAAC TGAA  
4440 4450 4460 4470 4480 4490 4500

1570 1580 1590 1600 1610 1620 1630 1640  
TGGGATAGGGTAAAAGGAAACCAAAGAGTGAGTGATATTAT-TGTGGAGGTGACAGACTACCCCTTCTGGGT  
| | | | | | | | | | | | | | | | | | | | | |  
TCAAAT-GAACAAATAGAAGAAACCGA-TGCTATCAAGATCTGTGCTGG-GA--GCATCTGAGTGAGGAGT  
4510 4520 4530 4540 4550 4560 4570

1650 1660 1670 1680 1690 1700 1710  
ACGTAGGGACAGACCTCCT-TCGGACTGTCTAAACTCCCCTTAGA-AGTCTCGTCAAGTTCCCGGACGAAG  
| | | | | | | | | | | | | | | | | | | | | |  
ACGT--GTGTAGCCGTCCTCTTGTA--GGCAAGCAGACGACATACAGAGTGCTGCTGTT--TGTACGGGG  
4580 4590 4600 4610 4620 4630

1720 1730 1740 1750 1760 1770  
AGGACAGAGGAGACACAGTCCGAAAAGT---TATTTTCCGGCAAAT-CCT-TTCCCTGTTTCGTGACACT  
| | | | | | | | | | | | | | | | | | | | | |  
AGG-CATGGGGCATGCAGTGTGCTCTGCCCCATGAAGGACTCAGATGACTATGCCAGCT--GTG-CA--  
4640 4650 4660 4670 4680 4690 4700

1780 1790 1800 1810 1820 1830 1840  
CCACCCCTTGTGGACACTTGAGTGTATCC--TTGCGCGGAAGGTGAGGTGGTAC--CCGT---CTGTAGG  
| | | | | | | | | | | | | | | | | | | | | |  
ACATCCC-TGT-GACAGGACGGCGGGACCATATGGACGGGATGCGTTGGTGG-ACTTCAGTGAACAGTA-T  
4710 4720 4730 4740 4750 4760 4770

1850 1860 1870 1880 1890 1900  
GGCGGGGAGACAGAGCCGCGGGGAGCTACGAGAATCGACT--CACAGGGCGCCCGG-GCTTC--GCAAAT  
| | | | | | | | | | | | | | | | | | | | | |  
GGCCCGAAACAGACCCTTACTTCA--TTC-AGGATCGCTTCTAAACAGCTTTGAGGAGCTACAGGCTGAG  
4780 4790 4800 4810 4820 4830 4840



CC plated to provide approximately 800 colonies per plate. The colonies  
CC were harvested and each pool used to prepare plasmid DNA for  
CC transfection into COS-7 cells. Transformants expressing biologically  
CC active cell surface G-CSFR were identified by screening for ability  
CC to bind 125-Iodine-G-CSF. Bacteria from a positive pool were plated  
CC and plasmids prepared. COS-7 cells were transfected and a single  
CC positive clone was identified and designated D-7. A glycerol stock  
CC of bacteria transformed with this G-CSFR cDNA clone in expression  
CC vector pCAV/NDT has been deposited as ATCC #68102.  
CC See also 011580.

S0 Sequence 2546 BP; 548 A; 844 C; 687 G; 467 T;

Initial Score = 138 Optimized Score = 967 Significance = 7.19  
Residue Identity = 47% Matches = 1196 Mismatches = 955  
Gaps = 342 Conservative Substitutions = 0

```

      10      20      30      40      50      60      70
ATGTCCATGAACTGCTGAGTGGATAAACAGCACGGGATATCTCTGTCTAAAGGAATATTACTACACCAGGAA
      | | | |      | | | | | | | | | | | | | | | |
      TGGACTGCA----GCTGGTTTCAGGAACCTCTCTTGACGA-GAA
          X          10          20          30

      80      90      100      110      120      130      140
AAGGACACATTTCGACAACAGGAAAGGAGCCTGTACAGAAAACACAGTGTCTGTGCATGTGA----CATT
| | | | | | | | | | | | | | | | | | | | | | | | | |
GAG-AGACCAAGGAGGCCAACGAGGGCTGGGCCAGAGGTGCCAACA-TG----GGGAAACTGAGGCTCGGC
40          50          60          70          80          90          100

      150      160      170      180      190      200
TCGCCATGG-GAAACAACGTGTACAACGTGGTGCATTGTGCTGCTG-----CTAGT--GGGCTGTGAGAA
| | | | | | | | | | | | | | | | | | | | | | | | | |
TCGGAAAGGTGAAGTAACTTGTCAA-----GATCACAAGCTGGTGAACATCAAGTTGGTGCTATGGCAA
110          120          130          140          150          160          170

      210      220      230      240      250      260
GG-TGGGAGCCGTGCAGAACT--CCTGTGAT-AACTG-TCAGCCTGGTACT-TTCTGCAGAAAATA-CAATC
| | | | | | | | | | | | | | | | | | | | | | | | | |
GGCTGGGAAAC-TGCAG-CCTGACTTGGGCTGCCCTGATCATCCTGCTGCTCCCGGAAGTCTGGAGGAGTG
180          190          200          210          220          230          240

270      280      290      300      310      320      330
CAGTCTGCAAGAG---CT--GCCCTCCAAGTACCTTCTCCAGCATAGGTGGACAGCCGAACGTAAACATC-T
| | | | | | | | | | | | | | | | | | | | | | | | | |
GGGGC-ACATCAGTGTCTCAGCCC-CCATCGTCCACCTGGGGGATCCCATCACAGCC-TCCTGCATCATCAA
250          260          270          280          290          300          310

      340      350      360      370      380      390      400
GCAGAGTGTGTGCAGGCTATTTCAAGTTCAAGAAAGTTTGTCTCTTACCCACAACGGGAGTGTGAGTGCA
| | | | | | | | | | | | | | | | | | | | | | | | | |
GCAGA---ACTGCA-GCCAT--CTGGACCCGGAGCCACAGATTCTGTGGAGACTGGGAGCAGAGCTTCAGCC
320          330          340          350          360          370

410      420      430      440      450      460      470
TTGAAGGATTCCATTGCTTGGGGCCACAGTGCACCAGATGTGAAAAGGACTGCAGGCCTGGCCAGGAGCT-A
| | | | | | | | | | | | | | | | | | | | | | | | | |
CGGGGCGAGGCAGCGTCTGTCTGATGGGACCCAG----GAATCTATCATCA-CCCTGCCC--CACCTCA
380          390          400          410          420          430          440

480      490      500      510      520      530      540
AC--GAAGCAGGGTTGCAAAACCTGTAGCTTGGGAAC--ATTTAATGACCAGAACGGTACTGCGCTGTGCG
| | | | | | | | | | | | | | | | | | | | | | | | | |
ACCACACTCAGGCCTTTCTCTCTGCTGCTGCT--GAACTGGGGCAACAGCCTGCA-GATCCTGGACCAAGTTG
450          460          470          480          490          500          510

550      560      570      580      590      600
ACCTCGACGAA--CTGCTCTCTAG-----ACCGAAGCTCT-CTG-CTTAAGACCGGGACCAAGGAGAAG
```

| | | | | | | | | | | | | | | | | | | | | |  
 A-GCTGCGCGCAGGCTACCTCCAGCCATACCCACAACCTCTCTGCGCTCATGAACCTCACAAC-CAGCAG  
 520 530 540 550 560 570 580  
 610 620 630 640 650 660 670  
 -GACGT-GGTGTGTGGACCCCTGTGGTGAAGTCTCTCCCA-GTACCACCATTTCTGTGA-----CTCCAG  
 | | | | | | | | | | | | | | | | | | | | | |  
 CCTCATCTGCCAGTGGGAGCCAGGACCTGAG--ACCCACCTACCCACCAGCTTCACTCTGAAGAGTTTCAAG  
 590 600 610 620 630 640 650  
 680 690 700 710 720 730  
 AG-GGAGG--AC---CAG----GAGGGCACTCCTTGCAGGTCCTTACCTTGTCTCTGGCGCTGACATCGGCT  
 | | | | | | | | | | | | | | | | | | | | | |  
 AGCCGCGGCAACTGTGAGACCAAGGGGACTCCATCCTGGAC--TGC-GTGCCCAAGG-ACGGGCAGAGCCA  
 660 670 680 690 700 710  
 740 750 760 770 780 790  
 TTGCTGCTGGCCCTGATCTTCACTTCTCTCTGTTCT--CTG----TGCTCAAATGGAT-CAGGAAAAAATT  
 | | | | | | | | | | | | | | | | | | | | | |  
 CTGCTGCATCCC---ACGCAAACAC-CTGCTGTTGTACCAGAATATGGGCATCTGGGTGCAGGCAGAGAAT  
 720 730 740 750 760 770 780  
 800 810 820 830 840 850 860  
 CCCCCACATATTCAAGCAACCATTTAAGAAGACCACTG-GAGCAGCTCAAGAGGAAGATG----CTTGTAGC  
 | | | | | | | | | | | | | | | | | | | | | |  
 GCGCTGGGGA--CCAG----CATGT--CCCCACAACCTGTGCTTGTATCCCATGGATGTTGTGAAACTGGAGC  
 790 800 810 820 830 840  
 870 880 890 900 910 920 930  
 TGCCGATGTCCACAGGAAGAAGAA--GGAG-GAGGAGGAGG-CTATGAGCTGTGATGTACTATCCTAGGAGA  
 | | | | | | | | | | | | | | | | | | | | | |  
 CCCCCATG-CTGC-GGACCATGGACCCAGCCCTGAAGCGGCCCTCCCAGGCAGG--CT-GCCTA-CAGC  
 850 860 870 880 890 900 910  
 940 950 960 970 980 990 1000  
 TGTG-TGGGCCGAAACCGAGAAGCA-CTAGGACCCACCATCCTGTGGAACAGCACAAGCAACCCACCACC  
 | | | | | | | | | | | | | | | | | | | | | |  
 TGTGCTGGG----AGCCATG--GCAGCCAGGCCTGCA-CATAAATCAGA--AGTGTGAGCTGCGCCACAAGC  
 920 930 940 950 960 970  
 1010 1020 1030 1040 1050 1060 1070  
 CTGTTCTTACACATCATCTAGATG---ATGTGTGGGCGCGCACCTCATCCAAGTCTCTTCTAAGCCTAACA  
 | | | | | | | | | | | | | | | | | | | | | |  
 CGCAGCGTGGAGA--AGCC-AGCTGGGCACTGGTGGGCCCCCTCC-CCTTGGAGGCCCTTC--AGTATGAGC  
 980 990 1000 1010 1020 1030 1040  
 1080 1090 1100 1110 1120 1130 1140  
 TATTTGTCTTTACCTTTTTTAAATCTTTTTTAAATTTAAATTTATGTGTGTG-AGTGTTTTGCTG---C  
 | | | | | | | | | | | | | | | | | | | | | |  
 TCTGCGGGCT--CCTCCCAGCCAGGCCCTACACCCTGCAGATACGCTGCATCCGCTGGGCCCTGCCTGGCCA  
 1050 1060 1070 1080 1090 1100 1110  
 1150 1160 1170 1180 1190 1200  
 CTGTATGCACACGTGTGTGTGTGTGTGTGTGACACTCCTGA--TGCCCTGAGGAGGTGAGA--AGAGAAAG  
 | | | | | | | | | | | | | | | | | | | | | |  
 CTGGAGCGACTGGAGCCCCAGCCTGGAGCTGAGA-ACTACCGAAGGGGCCCCCACTGTGAGCTGGACACAT  
 1120 1130 1140 1150 1160 1170 1180  
 1210 1220 1230 1240 1250 1260 1270  
 GGTGGTTCCATAAG-AACTGGAGTTATGGATGGCTGTG-AGCCGGNNNGATAG---GT----CGGGACGGGA  
 | | | | | | | | | | | | | | | | | | | | | |  
 GG-TGGCGGCAGAGGCAGCTGGACCCAGGA---CAGTGACAGCTGTTCTGGAAGCCAGTGCCCTGGAGGAA  
 1190 1200 1210 1220 1230 1240 1250  
 1280 1290 1300 1310 1320 1330  
 GAC--CTGTC--TTCCTATTTTAACTGCACTGTATAATAAAAAAAAAAATCATATTTCCCAATTCTACACA

[illegible]







550 560 570 580 590 600  
ACCTGACGAA--CTGCTCTCTAG-----ACGGAAGGTCT-GTG-CTTAAGACCGGACCACGGAGAAG  
| | | | | | | | | | | | | | | | | | | | | |  
A-GCTGCGCGCAGGCTACCCTCCAGCCATACCCACAACCTCTCTGCCTCATGAACCTCACAAC-CAGCAG  
520 530 540 550 560 570 580

610 620 630 640 650 660 670  
-GACGT-GGTGTGTGGACCCCTGTGGTGAGCTTCTCTCCCA-GTACCACCATTTCTGTGA-----CTCCAG  
| | | | | | | | | | | | | | | | | | | | | |  
CCTCATCTGCCAGTGGGAGCCAGGACCTGAG--ACCCACCTACCCACCAGCTTCACTCTGAAGAGTTTCAAG  
590 600 610 620 630 640 650

680 690 700 710 720 730  
AG-GGAGG--AC---CAG---GAGGGCACTCCTTGCGAGGTCTTACCTTGTTCCTGGCGCTGACATCGGCT  
| | | | | | | | | | | | | | | | | | | | | |  
AGCCGCGGCAACTGTGAGACCAAGGGGACTCCATCCTGGAC--TGC-GTGCCCAAGG-ACGGGCAGAGCCA  
660 670 680 690 700 710

740 750 760 770 780 790  
TTGCTGCTGGCCCTGATCTTCATTACTCTCCTGTTCT--CTG----TGCTCAAATGGAT-CAGGAAAAAATT  
| | | | | | | | | | | | | | | | | | | | | |  
CTGCTGCATCCC---ACGCAAACAC-CTGCTGTTGTACCAGAATATGGGCATCTGGGTGCAGGCAGAGAAT  
720 730 740 750 760 770 780

800 810 820 830 840 850 860  
CCCCACATATTCAAGCAACCATTTAAGAAGACCACTG-GAGCAGCTCAAGAGGAAGATG---CTGTAGC  
| | | | | | | | | | | | | | | | | | | | | |  
GCGCTGGGGA--CCAG---CATGT--CCCCAACAAGTGTCTTGATCCCATGGATGTTGTGAAACTGGAGC  
790 800 810 820 830 840

870 880 890 900 910 920 930  
TGCCGATGTCCACAGGAAGAAGAA--GGAG-GAGGAGGAGG-CTATGAGCTGTGATGTACTATCCTAGGAGA  
| | | | | | | | | | | | | | | | | | | | | |  
CCCCCATG-CTGC-GGACCATGGACCCAGCCCTGAAGCGGCCCTCCCCAGGCAGG--CT-GCCTA-CAGC  
850 860 870 880 890 900 910

940 950 960 970 980 990 1000  
TGTG-TGGGCCGAAACCGAGAAGCA-CTAGGACCCACCATCCTGTGGAACAGCACAAGCAACCCACCACC  
| | | | | | | | | | | | | | | | | | | | | |  
TGTGCTGGG---AGCCATG--GCAGCCAGGCCTGCA-CATAAATCAGA--AGTGTGAGCTGCCGCACAAGC  
920 930 940 950 960 970

1010 1020 1030 1040 1050 1060 1070  
CTGTTCTTACACATCATCCTAGATG---ATGTGTGGGCGCGCACCTCATCCAAGTCTCTTCTAACGCTAACA  
| | | | | | | | | | | | | | | | | | | | | |  
CGCAGCGTGGAGA--AGCC-AGCTGGGCACTGGTGGGCCCCCTCC-CCTTGAGGCCCTTC--AGTATGAGC  
980 990 1000 1010 1020 1030 1040

1080 1090 1100 1110 1120 1130 1140  
TATTTGTCTTTACCTTTTTTAAATCTTTTTTAAATTTAAATTTATGTGTGTG-AGTGTTTTGCTG---C  
| | | | | | | | | | | | | | | | | | | | | |  
TCTGCGGGCT--CCTCCCAGCCACGGCCTACACCCTGCAGATACGCTGCATCCGCTGGCCCCTGCCTGGCCA  
1050 1060 1070 1080 1090 1100 1110

1150 1160 1170 1180 1190 1200  
CTGTATGCACAGTGTGTGTGTGTGTGTGTGACACTCCTGA--TGCCTGAGGAGGTGAGA--AGAGAAAG  
| | | | | | | | | | | | | | | | | | | | | |  
CTGGAGCGACTGGAGCCCCAGCCTGGAGCTGAGA-ACTACCGAAGCGGCCCCCACTGTGAGACTGGACACAT  
1120 1130 1140 1150 1160 1170 1180

1210 1220 1230 1240 1250 1260 1270  
GGTTGGTTCCATAAG-AACTGGAGTTATGGATGGCTGTG-AGCCGGNNGATAG---GT----CGGGACGGA  
| | | | | | | | | | | | | | | | | | | | | |  
GG-TGGCGGCAGAGGCAGCTGGACCCAGGA---CAGTGCAGCTGTTCTGGAAGCCAGTGGCCCTGGAGGAA  
1190 1200 1210 1220 1230 1240 1250

[illegible]

2010 2020 2030 2040 2050 2060 2070

GCCTTGACTCCGGTACTAATTCTCCCTGCCGGCCCCCGTAAGCATA-ACGCGGCG-ATCTCCACTTTAAGAA  
| | | ||| | | ||||| | | | | | | | | | | | | | | | | |  
G--TACAGTCC-----TCACCCTGATG--ACCTTGACCCAGAGGGGTCGGAGGTACAC---ATCAT  
2000 2010 2020 2030 2040 2050

2080 2090 2100 2110 2120 2130 2140

CCTGGCGCGGTTCTGCCTGGTCTCGCT-TTCGTAACGGTTCTTACAAAGTAATTAGTTCTTGCTTTTCAGC  
||| | | ||||| ||| | | ||| | | | | | | | | | | | | | | | | |  
CCTGGGCGCTGTTGGGCC---TCCTGCTGTTGCTCACCTGCCTCTGTGGAAGTGCCTGGCTCT--GTTGCAGC  
2060 2070 2080 2090 2100 2110

2150 2160 2170 2180 2190 2200 2210

CTCCAAGCTTCTG-CTAGTCTATGGC--AGCATCAAGGCTGGTATTGCT-ACGGCTGAC--CGTAC--GC  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
C-CCAACAGGAAGAATCCCCTCTGGCCAAGTGTG---CCAGACCCAGCTCACAGCAGCCTGGGCTCCTGGG  
2120 2130 2140 2150 2160 2170 2180

2220 2230 2240 2250 2260 2270 2280

CGC-CGCAATAAGGGTACTGGGCGGCCCGTGAAGGCCCTTTGGTTTCA-GAAACCCAAGGCCCCCCCTATA  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
TGCCCAACATCATGG-AGGAGGATGCCTTCCAGCTGCCCGGCGCTTGGCAGCCACCCATCACCAGCTCA--  
2190 2200 2210 2220 2230 2240 2250

2290 2300 2310 2320 2330 2340

CCAACGTT--TCGACTTTG-ATTCTTGCCGGTACGTGGTGGTGGGTCCTTAGCTCTTTCTCGATAGTTAGA  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
-CAGTGCTGGAGGAGGATGAAAAGAAGCCGGT--GCCCTGG-GAGTCCCATACAGCTCAGAGACCTGTGGC  
2260 2270 2280 2290 2300 2310 2320

2350  
C  
|  
CTCCCCACTCTGGTCCAGACCTATGTGCTCCAGGGGGACCCAAGAGCAGTT  
X 2330 2340 2350 2360 2370

5. ELLIS-012-FIG2AB.SEQ (1-2350)

Q13856 Human GCSF receptor gene in pHQ3/pHG12.

ID Q13856 standard; DNA; 2942 BP.

AC Q13856;

DT 08-JAN-1992 (first entry)

DE Human GCSF receptor gene in pHQ3/pHG12.

KW Granulocyte colony stimulating factor; ss.

OS Homo Sapiens.

FH	Key	Location/Qualifiers
----	-----	---------------------

FT	sig_peptide	169..237
----	-------------	----------

FT /\*tag= a

FT Rat peptide 238..2676

FT /\*tag= b

PN W09114776-A.

PD 03-OCT-1991.

PF 22-MAR-1991; J00375.

PR 23-MAR-1990; JP-074539.

PR 03-JUL-1990; JP-176629.

PA (OSAB-) OSAKA BIOSCIENCE IN.

PI Nagata S, Fukunaga R;

DR WFI; 91-310576/42.

DR P-PSDB; R14255.

PT DNA encoding granulocyte colony stimulating factor receptor - for

PT recombinant prodn. of GCSF receptor useful in therapy and

PT research.

PS Claim 1; Fig 8; 99pp; Japanese.

CC The sequence was obtd. from a cDNA library prepd. from human

CC histocytic lymphoma H977 cells using DNA from the corresponding

CC murine gene (see Q13855). The genes can be used to produce  
CC recombinant receptors for use in research and for diagnostic assays.  
CC See also Q13857 and Q13858.

SQ Sequence 2942 BP; 611 A; 993 C; 796 G; 542 T;

Initial Score = 138 Optimized Score = 971 Significance = 7.19  
Residue Identity = 48% Matches = 1203 Mismatches = 951  
Gaps = 345 Conservative Substitutions = 0

[illegible]

GCTCATCTGCCAGTGGGACGCAGGACCCTGAG--ACCCACCTACCCCACCAGCTTCACCTCTGAAGAGTTTCAA  
590           600           610           620           630           640           650

        680                 690             700             710             720             730  
GAG--GGAG-GAC---CAG----GAGGGCACTCCTTGCGAGGTCTTACCTTGTTCCTGGCGCTGACATCGGC  
||| || | || ||| |||| |||| | | | | | | | | | | | | | |  
GAGCGGGGCAACTGTGAGACCCAAGGGGACTCCATCCTGGAC--TGC-GTCCCCAAGG-ACGGGCAGAGCC  
660           670           680           690           700           710           720

        740             750             760             770                 780             790  
TTTGCTGCTGGCCCTGATCTTCATTACTCTCCTGTTCT--CTG----TGCTCAAATGGAT-CAGGAAAAAAAT  
||||| || | | ||| |||| | | | | | | | | | | | | | |  
ACTGCTGCATCCC---ACGCAAACAC-CTGCTGTTGTACCAGAATATGGGCATCTGGGTGCAGGCAGAGAA  
730                 740             750             760             770             780

800           810           820           830           840           850           860  
TCCCCACATATTCAAGCAACCATTTAAGAAGACCACTG-GAGCAGCTCAAGAGGAAGATG---CTTGTAG  
| | | | | | | | ||| | ||||| | | | | ||| | | | | | |  
TGCGCTGGGGA--CCAG---CATGT--CCCCACAACGTGTCTTGATCCCATGGATGTTGTGAAACTGGAG  
790           800           810           820           830           840           850

        870             880             890             900             910             920             930  
CTGCCGATGTCCACAGGAAGAAGAA--GGAG-GAGGAGGAGG-CTATGAGCTGTGATGTACTATCCTAGGAG  
| || ||| | | ||| ||| || | ||| | | | | | | | | | ||| ||  
CCCCCATG-CTGC-GGACCATGGACCCAGCCCTGAAGCGGCCCCCTCCCAGGCAGG--CT-GCCTA-CAG  
860           870           880           890           900           910

        940             950             960             970             980             990           1000  
ATGTG-TGGGCCGAAACCGAGAAGCA-CTAGGACCCACCATCCTGTGGAACAGCACAAAGCAACCCACCAC  
|||| |||| | || | ||| |||| | || ||| | | | | | | | | | ||| ||  
CTGTGCTGGG---AGCCATG--GCAGCCAGGCCTGCA-CATAATCAGA--AGTGTGAGCTGCGCCACAAG  
920           930           940           950           960           970           980

        1010            1020            1030            1040            1050            1060            1070  
CCTGTTCTTACACATCATCTAGATG--ATGTGTGGGCGCGCACCTCATCCAAGTCTCTTCTAACGTAAC  
|| | | | | | ||| || | ||||| | | | | | | | | | ||| ||| | | |  
CCGCAGCGTGAGAA--AGCC-AGCTGGGCACTGGTGGGCCCCCTCC-CCTTGAGAGCCCTTC--AGTATGAG  
990           1000           1010           1020           1030           1040

        1080            1090            1100            1110            1120            1130  
ATATTTGTCTTTACCTTTTTTAAATCTTTTTTTAAATTTAAATTTATGTGTGTG-AGTGTTTGCGCTG---  
| | | | | ||| | | | | | | | | | | | | | | | | | | |||||  
CTCTGCGGGCT--CCTCCAGCCACGGCCTACACCTGCAGATACGCTGCATCGCTGCCCCCTGCGCTGGCC  
1050           1060           1070           1080           1090           1100           1110

1140           1150           1160           1170           1180           1190           1200  
CCTGTATGCACACGTGTGTGTGTGTGTGTGTGTGACACTCTGA--TGCCTGAGGAGGTGAGA--AGAGAAA  
||| | || | | | | ||| ||| ||| ||| ||| | | | | | | | | | |  
ACTGGAGCGACTGGAGCCCCAGCCTGGAGCTGAGA-ACTACCGAACGGGCCCCCACTGTCAGACTGGACACA  
1120           1130           1140           1150           1160           1170           1180

1210           1220           1230           1240           1250           1260           1270  
GGGTTGGTTCATAAG-AACTGGAGTTATGGATGGCTGTG-AGCCGGNNNGATAG---GT---CGGGACGG  
|| ||| || | | | |||| | || | ||| ||| | | | | | | | | | |  
TGG-TGGCGGCAGAGGCAGCTGGACCCAGGA---CAGTGCAGCTGTTCTGGAAGCCAGTGCCCTTGAGGA  
1190           1200           1210           1220           1230           1240           1250

        1280             1290             1300             1310             1320             1330  
AGAC--CTGTC--TTCTATTTTAAGCTGACTGTATAATAAAAAAAAAAATGATATTCGGGAATTGTAGAGA  
||| | | | | || | ||| ||| ||| | | | | | | | | | | | |  
AGACAGCGGACGGATCCAAGTT-ATGTG---GTTTCTTGA-----GACCT-CAGGCCAGGCTGGGG  
1260           1270           1280           1290           1300           1310

1340           1350           1360           1370           1380           1390           1400  
TTGTCCTGACACCTTCTA-GTTAATGATCTAAGAGGAATTGTTGATACGTAGTATACTGTATATGTGTATG

[illegible]

```

TCTGGGCGCTGTTTCGGCC---TCCTGCTGTTGCTCACCTGCTCTGTGGAACTGGCTGCTCT--GTTGCAG
2060      2070      2080      2090      2100      2110      2120

2150      2160      2170      2180      2190      2200
CCTCCAAGCTTCTG-CTAGTCTATGGC--AGCATCAAGGCTGGTATTTGCT-ACGGCTGAC--CGCTAC--G
|| ||||      | | || |||| || ||      | |      ||| || || || || || || ||
CC-CCAACAGGAAGAATCCCTCTGGCCAAGTGC----CCAGACCCAGCTCACAGCAGCCTGGGCTCCTGG
      2130      2140      2150      2160      2170      2180

2210      2220      2230      2240      2250      2260      2270
CCGC-CGCAATAAGGGTACTGGGCGGCCCGTGAAGGCCCTTTGGTTTCA-GAAACCCAAGGCCCCCTCAT
|| | |||| | || | || ||| | ||||      | || | |||| || || |||
GTGCCACAAATCATGG-AGGAGGATGCCTTCCAGCTGCCCGGCCTTGGCACGCCACCCATCACCAAGCTCA-
2190      2200      2210      2220      2230      2240      2250

2280      2290      2300      2310      2320      2330      2340
ACCAACGTT--TCGACTTTG-ATTCTTGCCGGTACGTGGTGGTGGTGCCTTAGCTCTTTCTCGATAGTTAG
|| ||      || || | ||||| | ||| || || || || || | || || ||
--CAGTGCTGGAGGAGGATGAAAAGAAGCCGGT--GCCCTGG-GAGTCCATAACAGCTCAGAGACCTGTGG
2260      2270      2280      2290      2300      2310      2320

X
AC
|
CCTCCCCACTCTGGTCCAGACCTATGTGCTCCAGGGGGACCCAAGAGCAGTT
2330      2340      2350      2360      2370

```

# 6. ELLIS-012-FIG2AB.SEQ (1-2350)

N61379 Sequence encoding porcine beta-follicle stimulin

```

ID N61379 standard; cDNA; 728 BP.
AC N61379;
DT 03-AUG-1992 (first entry)
DE Sequence encoding porcine beta-follicle stimulating hormone (FSH).
KW Superovulation therapy; hypophyseal disorder; gonadal regression;
KW infertility; ss.
OS Pig.
FH Key Location/Qualifiers
FT CDS 1..54
FT /*tag= a
FT transit_peptide 55..108
FT /*tag= b
FT mat_peptide 109..444
FT /*tag= c
FT CDS 445..726
FT /*tag= d
PN FR2565599-A.
PD 13-DEC-1985.
PF 07-JUN-1985; 508647.
PR 08-JUN-1984; US-618466.
PA (INTE-) INTEGRATED GENETICS.
PI Beck AK;
DR WPI; 86-030537/05.
DR P-PSDB; P61785.
PT New DNA coding for porcine beta-follicle stimulating hormone -
PT useful for raising antibodies, inducing ovulation etc., and new
PT expression vectors
PS Disclosure; Page 3; 14pp; French.
CC Total RNA is extracted from pig hypophyseal glands and used to
CC construct a library of cDNA. The library was screened using two
CC oligonucleotide probes designated PF55 and PF434. These were
CC ligated to give the complete sequence including the untranslated
CC flanking regions. This sequence has been inserted into pBR322 and
CC deposited as NRRL B-15793. The final vector is ppFSH.
SQ Sequence 728 BP; 187 A; 184 C; 147 G; 190 T;

```



Initial Score = 137 Optimized Score = 321 Significance = 7.13  
Residue Identity = 49% Matches = 389 Mismatches = 303  
Gaps = 94 Conservative Substitutions = 0

```

10      20      30      40      50      60      70
TGAAGTCTGAGTGGATAACAGCACGGGATATCTCTCTCTAAAGGAATATTAC-TACACCAGGAAAAGGAC
                                     ||| ||| || | ||
                                     GTACTTTCAC--GGTCTCGTAC
                                     X      10      20

80      90      100     110     120     130     140
AC--ATTGCACAACAGGAAAGGAGCCTGTACAGAA--AAACCACAG-TGTCCTG-TGCATGTGACATTTTCGC
||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
ACCAGCTCCTTAATTGTTTGGTTTCCACCCCAAGATGAAGTCGCTGCAGTTTGTCTCCTATTCTGTTGC--
      30      40      50      60      70      80      90

150     160     170     180     190     200     210
CATGGGAA-ACAACTGTTACAACGTGGTGGTCA-TTGTGCTGCTGCTAGTGGGCTGT-GAGAAGGTGGGAGC
||| || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
--TGGAAAGCCATCTGCTGCAA--TAGCTGTGAGCTGACCAACATCACCATCACAGTGGAGAAAG-AGGAG-
      100     110     120     130     140     150

220     230     240     250     260     270     280
CGTGCAGAACTCCTGTGATAACTGTCA--GCCTGGTACT-TTCTGCAGAAAATACAATCCAGTCTGCAAGAG
||  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
--TG--TAACTTCTG-CATAAGCATCAACACCACGTGGTGTGCTG--GCTATTGC--TACACCCGG---GAC
      160     170     180     190     200     210

290     300     310     320     330     340
CTGCCCTCCAAGTACCTTCTCCAGCATAGGTGGACAGCC-GAA---CTGTAACATC-TGCAGAGTGTGTGCA
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
CTGGTATACAAGGAC-----CCAGCCAGGCCCAACATCCAGAAAACATGTACCTTCAAGGAGCTGGTGTACG
      220     230     240     250     260     270     280

350     360     370     380     390     400     410     420
GGCTATTTTCAGGTTCAAGAAGTTTGTCTCTCTACCCACAACGCGGAGTGTGAGTGCATTGAAGGATTCCAT
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
AGACCGTGAAAGTACCTG--GCTGTGCT-CACCATGCA-GACTCCCTGTATACGT--AT-CCAGTA-GCCAC
      290     300     310     320     330     340

430     440     450     460     470     480     490
TGCTTGGGGCCACAGTGCACCAGATGTGAAAGGACTGCA-GGCCTG-GCCAGGAGCTAAC-GAAGCAGGGT
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
CGAAT---GTCACTGTG-GCAAG-TGTGACAGTGACAGTACTGACTGCACCGTGAGAGGCCCTGGGGCCAGC
      350     360     370     380     390     400     410

500     510     520     530     540     550
TGCAAAACCTGTAGCTTGGGAACATTTAATGACCAGAACGGTAC-TGGCGT-CTGTGACCCTGGAC-GAAC
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
TACTGCTCCTTCAG--TGAAATGAAAGAATAAAGAGCAGTGGACATTTTCATGCTTCCTACCCTTGTCTGAAG
      420     430     440     450     460     470     480

560     570     580     590     600     610     620
TGCTCTCTAGACG--GAAG--GTCTGTGCTTAAGACCGGGACCACGGAGAAGGACGTGGTGTGTGGACCCC
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
GAC---CAAGACGTCCAAGAAGTTTGTGTGTA--CATGTGCCCAGGCTGCA-AAC-CACTATGAGAGACCCC
      490     500     510     520     530     540

630     640     650     660     670     680     690
TGTGGTGAGCTTCTCTCC-CAGTACCACCATTTCTGTGACTCCAGAGGG--AGGACCAGGAGGGCA-CTCCT
||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
ACTGAT-CCCTGCTGTCTGTGGAGAGGAGCTCCAGGAATGCAGAGTGTAGGGCCTCAGTCCCATCACCA
      550     560     570     580     590     600     610     620

700     710     720     730     740     750     760
```



80           90           100           110           120           130           140  
AC--ATTCGACAACAGGAAGAAGGAGCCTGTGCACAGA--AAACCACAG-TGTCCTG-TGCATGTGACATTTCGC  
||       ||       ||       |       ||       ||       ||       ||       ||       ||       ||  
ACCAGCTCCTTAATTGTTTGGTTTTCCACCCAAGATGAAGTCGCTGCAGTTTTGCTTCCTATTCTGTTGC--  
          30          40          50          60          70          80          90

150          160          170          180          190          200          210  
CATGGGAA-ACAAC TGTACAAC TGGTGCTCA-TTGTGCTGCTGCTAGTGGGCTGT-GAGAAGGTGGGAGC  
   || | || | || | || | || | || | || | || | || | || | || | || | || | || | || | || |  
--TGGAAGGCCATCTGCTGCAA--TAGCTGTGAGCTGACCAACATCACCATCACAGTGGAGAAAG-AGGAG-  
          100         110         120         130         140         150

220          230          240          250          260          270          280  
CGTGCAGAACTCCTGTGATAACTGTCA--GCCTGGTACT-TTCTGCAGAAAATACAATCCAGTCTGCAAGAG  
   ||       |||| | || | || | || | || | || | || | || | || | || | || | || | || | || |  
--TG--TAACTTCTG-CATAAGCATCAACACCAGTGGTGTGCTG--GCTATTGC--TACACCCGG---GAC  
          160         170         180         190         200         210

290          300          310          320          330          340  
CTGCCCTCCAAGTACCTTCTCCAGCATAGTGGACAGCC-GAA---CTGTAACATC-TGCAGAGTGTGTGCA  
|| | | || | || | || | || | || | || | || | || | || | || | || | || | || | || | || |  
CTGTATACAAGGAC-----CCAGCCAGGCCCAACATCCAGAAAACATGTACCTTCAAGGAGCTGGTGTACG  
220         230         240         250         260         270         280

350          360          370          380          390          400          410          420  
GGCTATTTTCAGGTTCAAGAAGTTTTGCTCCTCTACCCACAACGCGGAGTGTGAGTGCATTGAAGGATTCCAT  
||       |       || | || | || | || | || | || | || | || | || | || | || | || | || |  
GGACCGTGAAAGTACCTG--GCTGTGCT-CACCATGCA-GACTCCCTGTATACGT--AT-CCAGTA-GCCAC  
290         300         310         320         330         340

430          440          450          460          470          480          490  
TGCTTGGGGCCACAGTGCACCAGATGTGAAAAGGACTGCA-GGCCTG-GCCAGGAGCTAAC-GAAGCAGGCT  
| |       || | || | || | || | || | || | || | || | || | || | || | || | || | || |  
CGAAT---GTCACTGTG-GCAAG-TGTGACAGTGACAGTACTGACTGCACCGTGAGAGGCTGGGGCCCAGC  
350         360         370         380         390         400         410

500          510          520          530          540          550  
TGCAAAACCTGTAGCTTGGGAACATTTAATGACCAGAACGGTAC-TGGCGT-CTGTCGACCCTGGAC-GAAC  
| |       || | || | || | || | || | || | || | || | || | || | || | || | || | || |  
TACTGCTCCTTCAG--TGAAATGAAAGAATAAAGAGCAGTGGACATTTTCATGCTTCCTACCTTGTCTGAAG  
420         430         440         450         460         470         480

560          570          580          590          600          610          620  
TGCTCTCTAGACG--GAAG--GTCTGTGCTTAAGACCGGACACGGAGAGGAGCTGGTGTGTGGACCCCC  
| |       |||| | || | || | || | || | || | || | || | || | || | || | || | || | || |  
GAC---CAAGACGTCCAAGAAGTTTGTGTGTA--CATGTGCCCAGGCTGCA-AAC-CACTATGAGAGACCCC  
490         500         510         520         530         540

630          640          650          660          670          680          690  
TGTGGTGAGCTTCTCTCC-CAGTACCACCATTTCTGTGACTCCAGAGGG--AGGACCAGGAGGGCA-CTCCT  
|| | || | || | || | || | || | || | || | || | || | || | || | || | || | || | || |  
ACTGAT-CCCTGCTGCTCTGTGGAGGAGGAGCTCCAGGAATGCAGAGTGCTAGGGCCTCAGTCCCATCACA  
550         560         570         580         590         600         610         620

700          710          720          730          740          750          760  
TGCAGGTCCTTACCTTGTTCCTGGCGCTGACATCGGCTTTGCTGCTGGCCCTGATCTT--CATTACTCTCCT  
||       |       || | || | || | || | || | || | || | || | || | || | || | || | || |  
CTCAACCCTGTATTTTGGGTCTGG---TTCCATAAG-TTTTATTCGGTCTTTTTTTTTTAAATTACTC-AA  
630         640         650         660         670         680

770          780          790          800          810          820          830  
G--TTCTCTGTGCTCAAATGGATCAGGAAAAAATCCCCACATATTCAAGCAACCATTTAAGAGACCACT  
| || | || | || | || | || | || | || | || | || | || | || | || | || | || | || | || |  
GAATTTTAT-TACATTTATAATTGTAGCAAGGAT--CATCACAA  
690         700         710         720                   x

## 8. ELLIS-012-FIG2AB.SEQ (1-2350)

Q03847 Porcine beta FSH subunit.

ID Q03847 standard; cDNA; 780 BP.  
 AC Q03847;  
 DT 24-AUG-1990 (first entry)  
 DE Porcine beta FSH subunit.  
 KW Luteinizing hormone; follicle stimulating hormone;  
 KW recombinant cDNA; alpha subunit; beta subunit; ungulate; ss.  
 OS Bos taurus.  
 FH Key Location/Qualifiers  
 FT CDS 107..496  
 FT /\*tag= a  
 FT /product=Porcine beta FSH  
 PN W09002757-A.  
 PD 22-MAR-1990.  
 PF 02-SEP-1988; 030949.  
 PR 02-SEP-1988; W0-U03049.  
 PA (INTE-) Integrated genetics.  
 PI Beck A, Bernstine E, Hsiung N, Kelton C, Lerner T, Reddy VB; Chappel SC.  
 DR WPI; 90-115954/15.  
 PT Biologically active ungulate LH and FSH- produced by recombinant methods.  
 PS Disclosure; Fig 10; 66pp; English.  
 CC LH and FSH comprises an alpha and a beta subunit, both subunits can be  
 CC synthesised in a single cell contg. an expression vector comprising  
 CC heterologous DNA encoding one subunit.  
 CC See also Q03843-Q03851.  
 SO Sequence 780 BP; 201 A; 195 C; 184 G; 200 T;

Initial Score = 137 Optimized Score = 339 Significance = 7.13  
 Residue Identity = 48% Matches = 405 Mismatches = 337  
 Gaps = 93 Conservative Substitutions = 0

```

      10      20      30      40      50      60
ATGTCCATGAAGTCTGAGTGGATAAACAGCACGGGATATCT-----CTGTCTAAAGGAATATTAC-TACAC
  ||  |||  |||||  | || ||  |||  ||  |||  |||  |||  |||
GAGTGGCTACCTGGATACGTA-TACAGGGAGTCTGCATGGTGAGCACAGCCA-AGTACTTTCAC
  X      10      20      30      40      50      60

```

```

      70      80      90     100     110     120     130
CAGGAAAAGGACAC--ATTGACAACAGGAAAGGAGCCTGTCACAGA--AAACCACAG-TGTCCTG-TGCAT
  ||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
--GGTCTCGTACACCAGCTCCTTAATTGTTTGGTTTCCACCCCAAGATGAAGTCGCTGCAGTTTGTCTTCT
  70      80      90     100     110     120     130

```

```

     140     150     160     170     180     190     200
GTGACATTTGCCATGGGAA-ACAACTGTTACAACGTGGTGGTCA-TTGTGCTGCTGCTAGTGGGCTGT-GA
  |  ||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
ATTCTGTTGC---TGGAAAGCCATCTGCTGCAA--TAGCTGTGAGCTGACCAACATCACCATCACAGTGGGA
  140     150     160     170     180     190

```

```

     210     220     230     240     250     260     270
GAAGGTGGGAGCCGTCAGAACTCCTGTGATAACTGTCA--GCCTGGTACT-TTCTGCAGAAAATACAATCC
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
GAAAG-AGGAG---TG--TAACCTTG-CATAAGCATCAACACCACGTGGTGTGCTG--GCTATTGC--TAC
200      210      220      230      240      250

```

```

     280     290     300     310     320     330
AGTCTGCAAGAGCTGCCCTCCAAGTACCTTCTCCAGCATAGGTGGACAGCC-GAA---CTGTAAACATC-TGC
  |  ||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
ACCCGG---GACCTGGTATACAAAGGAC-----CCAGCCAGGCCCAACATCCAGAAAACATGTACCTTCAAGG
240      270      280      290      300      310      320

```

```

340      350      360      370      380      390      400
AGAGTGTGTGACGGCTATTTCAAGTTCAGGTTCAAGAAGTTTTGCTCCTCTACCCACAACGCGGAGTGTGAGTGCATT
|| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
AGCTGGTGTACGAGACCGTGAAAGTACCTG--GCTGTGCT-CACCATGCA-GACTCCCTGTATACGTATCCA
      330      340      350      360      370      380      390

410      420      430      440      450      460      470
GAAGGATTCCATTG-CTTGGGGCCACAGTGCACCAGATGTGA-AAAGGACTGCAGGCCTGGCCAGGAGCTAA
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
GTAGCCACCGAATGTCACTGTGGCA-AGTGTGACAG-TGACAGTACTGACTGCA--CCGTGAGAGG-CCT--
      400      410      420      430      440      450

480      490      500      510      520      530      540
CGAAGCAGGGTTGCAAAACCTGTAGCTTGGGAACATTTAATGACCAGAACGGTAC-TGGCGT-CTGTGACC
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
-GGGGCCCAGTACTGCTCCTTCAG--TGAAATGAAGAATAAAGAGCAGTGGACATTTTCATGCTTCTACC
      460      470      480      490      500      510      520

550      560      570      580      590      600      610
CTGGAC-GAAGTGTCTCTAGACG--GAAG--GTCTGTGCTTAAGACGGGACCACGGAGAAGGACGTGGTG
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
CTTGCTGAAGGAC---CAAGACGTCCAAGAAGTTTGTGTGTA--CATGTGCCAGGCTGCAAACCACTATG
      530      540      550      560      570      580      590

620      630      640      650      660      670      680
TGTTGACCCCCCTGTGGTGAETTTCTCTCCAGTACCACCATTTCTGTGACTCCAGAGGG--AGGACCAGGAG
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
AGAGACCCCACTGATCCCTGC-TGTC-CTGTGGAGGAGGAGCTCCAGGAATGCAGAGTGCTAGGGCCTCAGT
      600      610      620      630      640      650      660

690      700      710      720      730      740      750
GGCA-CTCCTTGCAGGTCTTACCTTGTTCCTGGCGCTGACATCGGCTTTGCT-GCTGGCCCTGATCTTCAT
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
CCCATCACCCTCAACCCTGATTTTGGGTCTGG---TTCCATAAGTTTATTCGGTCTTTTTTTTTTAAAT
      670      680      690      700      710      720      730

760      770      780      790      800      X      810      820
TACTCTCTG--TTCTCTGTGCTCAAATGGATCAGGAAAAAATCCCCACATATTCAAGCAACCATTTAAG
|||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
TACTC-AATGAATTTTAT-TACATTTATAATTGTAGCAAGGATCATCACA
      740      750      760      770      780

830      840      850
AAGACCACTGGAGCAGCTCAAGAGGAAGA

```

9. ELLIS-012-FIG2AB.SEQ (1-2350)

Q28758 Partial sequence of tumour suppressor gene U10.

ID Q28758 standard; DNA; 4328 BP.

AC Q28758;

DT 25-FEB-1993 (first entry)

DE Partial sequence of tumour suppressor gene U10.

KW CaN19; tumour suppressor gene; cancer; therapy; ss.

OS Homo sapiens.

PN W09215602-A.

PD 17-SEP-1992.

PF 28-FEB-1992; U01624.

PR 28-FEB-1991; US-662216.

PA (DAND ) DANA FABER CANCER INST INC.

PI Sager R

DR WPI; 92-331663/40.

PT Diagnoses and treatment of cancer - using candidate tumor suppressor

PT genes or the corresp. antibodies.

PS Claim 30; Page 38-41; 56pp; English

CC An adaptation of the subtractive hybridization technique was used  
 CC which utilizes a biotinylation-based subtraction procedure instead  
 CC of hydroxyapatite as previously used. In this procedure, a single  
 CC strand phagenid cDNA library from normal cell polyA+ mRNA is  
 CC hybridized with excess biotinylated tumor polyA+ mRNA, and the  
 CC resulting double stranded sequences are removed by binding to  
 CC streptavidin. The remaining single-stranded phagenid cDNAs are  
 CC converted to double-stranded form and used to transform bacterial  
 CC host cells. The resulting subtracted cDNA library is differentially  
 CC screened with total cDNA from normal and tumor cells. This method  
 CC produced some 20 additional cloned cDNAs. Also found by this  
 CC method were several genes which, on the basis of the partial DNA  
 CC sequences appear to be novel sequences not previously entered  
 CC into GENBANK. The portion of the cDNAs so sequenced represents  
 CC part of the coding region and/or part of the 3' untranslated region  
 CC of each cDNA (see Q28749-58).

58 Sequence 4328 BP; 1236 A; 970 C; 912 G; 1210 T;

Initial Score = 135 Optimized Score = 953 Significance = 7.01  
 Residue Identity = 46% Matches = 1140 Mismatches = 1017  
 Gaps = 279 Conservative Substitutions = 0

```

                                X      10      20
                                ATGTCATGAAGTCTGAGTGG
                                | | ||||| | | |
CAGTTATGTTCTGTTTCGTTATTGGTACCAAACTCTTGCCAGATAACCAGTTTCATGAAGTGT---TGT
1990      2000      2010      2020      2030      2040      2050

      30      40      50      60      70      80      90
ATAAACAGCAGCGGATATCTCTGTCTAAAGGAATATTACTACACCAGGAAAAGGACACATTGAC--AACAG
|| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
AT-GGCAGCCCATGTTCTCTAATGCCACTGCTCTGTT-TTA-AAAGCTCAGAGG-CAATTTTACATATCAG
2060      2070      2080      2090      2100      2110      2120

      100      110      120      130      140      150
GAAAGGAGCCTGTCACAGAAAACACAG-TGTCCTG-TGCATGTGACATTTGCCATGGGAAACAAGT---
|| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
TAATTG---TTTTTATA-ATTTGCATGTTTTTCATGAAACAT-TGCTATGCATTTATTAGGAAAACTGAAT
2130      2140      2150      2160      2170      2180      2190

160      170      180      190      200      210      220
TTACAACGTGGTGTC---ATTGTGCTGCT-GCTAGTGGGCTGTGAGAAGGTGGGAGCCGTGCAGAACTCCT
|| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
TTCCAACAGGTGAAGTGAAGTTATTTTAACTATTATAC-ATAATCA-GAAAGATCC-TGC--CTCTACG
2200      2210      2220      2230      2240      2250

230      240      250      260      270      280      290
GTGATAACTGTCAGCCTGGTACTTTTTCAGAGAAATACAATCCAGTCTGCAAGAGCTGCCCTCCAAGTACC-
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
GAATTAGC--TAAACCTAAAAATGTTTGCATTAA--TGAATAAATTCTT-----CTGCATTCCTTGGCCCA
2260      2270      2280      2290      2300      2310      2320

300      310      320      330      340      350      360
-TTCTCCAGCATAGGTGGACAGCCGAAGTGAACATCTGCAGAGTGTGTGACAGGCTATTTCAAGTTCAAGAA
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
GTTCTGGAG--TTGGTGACCTTTATCACAATTATAT-TTTAG--GCGGCCAGTGAAGTGTGCTTC-AGAA
2330      2340      2350      2360      2370      2380

370      380      390      400      410      420      430
GT---TTTGCTC--CTCT-ACCCACAACGCGGAGTG---TGAGTGCA--TTGAAGGATTC-CATTGCTTGGGG
|| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
GTCCATAGCCAGCTCTGAAGTTTCTCGATAAATGCCATCAGTTCACCTTTAAAGACACACATTCTTTG--
2390      2400      2410      2420      2430      2440      2450

      440      450      460      470      480      490
CCAGAGCTGCAAGACATCTGAAAGAGTGGAGCCCTGCGGAGGAGCTAAGCAAGCAGCC---TTGCAAAAGC

```

```

| | | | | | | | | | | | | | | | | | | | | |
--AAA-TCCACCCAGTGTAA---AAAGCA-ACCTGGAAATTTAC-ACATTAGCATTGTACTTTCTAGCCC
2460      2470      2480      2490      2500      2510

500      510      520      530      540      550      560
TG TAGCTTGGGAACATTTAATGACCAGAACGGTACTG-GCGTCTGTCGACCCTG-GACGAAC-TGCTCTCTA
| | | | | | | | | | | | | | | | | | | | | |
--TAATTTGTGAGGTGTCAGCTATCATTATATTCTGCATGTATGTATAACCTGTTGTGAACAATCATACTT
2520      2530      2540      2550      2560      2570      2580

570      580      590      600      610      620      630
GACGGAAGGTCTG-TG-CTTAAGACCGGGACCACGGAGAAGGAC-GTGGTGTGTGGACCCCTGTGGTGAGC
| | | | | | | | | | | | | | | | | | | | | |
AACAAACTACTGATGGTTTATGAC---AACGTAGGTAACACAGTTCATTCTGTTCC-----AGGTTATA
2590      2600      2610      2620      2630      2640      2650

640      650      660      670      680      690      700
TTCTCTCCAGTACCACCATTTCTGTGACTCCAGAGG--GAGGACCAGGAGGCACTCCTTGCAGGTCCTTA
| | | | | | | | | | | | | | | | | | | | | |
TAAAACTGCATTTCTGAATTTGGTTAAAACTAAGGATGATGGATTGAAACAGTCTTTTAAATTAGTTT
2660      2670      2680      2690      2700      2710      2720

710      720      730      740      750      760      770
CCTTGTTCCTGGCG--CTGACATCGGCTTTGCTGCTGGCCCTGA-TC-TTCATTACTCTCCTGTTCTCTGT-
| | | | | | | | | | | | | | | | | | | | | |
ATATGCTTTAGGTGTTTTGGAATTTGCCCTTCTTGAACCTCCTGAGTCACACAGAAAGCAACTGTACACAGTA
2730      2740      2750      2760      2770      2780      2790

780      790      800      810      820      830      840
-GCTCAATGGATCAGGAAA---AAATTCCTCCACATATCAAGCAACCATTTAAGAAGACCACTG-GAGC
| | | | | | | | | | | | | | | | | | | | | |
GAATTCTGTGGCGCAGACCATGCTGTATTAACACATCACTTGCTGTTTCTACTGAGTGTACCACTGCCTTC
2800      2810      2820      2830      2840      2850      2860

850      860      870      880      890      900
AGCTCAAG---AGGAAGATGCTTGTAGCTGCCGA-TGTC---CACAGGAAGAAGA-AGGAGGAGGAGGAGGC
| | | | | | | | | | | | | | | | | | | | | |
CCTTCTAGCCAGGAGAATG-TTTACTCAGTTTAGTGCTTTGTATTTCTATAATACCAACAGGA--ATGG
2870      2880      2890      2900      2910      2920      2930

910      920      930      940      950      960      970
TA-TGA-GCTGTGATGTACT--ATCCTAGGAGATGTGTGGCCGAAACCGAGAA-GCACTAGGACCCCA-CC
| | | | | | | | | | | | | | | | | | | | | |
TAGTCACACTGTCTGAAATTGAATCT-GTCCATCTGT---TTATAATCAAGAACATATCAGAAATATATAG
2940      2950      2960      2970      2980      2990      3000

980      990      1000      1010      1020      1030
ATCCTGTGGAACA-GCACAAGCAACCC-CACCACCTGTT-----CTTACACATCATCCT--AGATGAT--G
| | | | | | | | | | | | | | | | | | | | | |
GTCCAGGTAATACTCCCAACATCCCACTTTTACTGTTTCAGGCCATCATATCATTCTTAAGCTACTTGG
3010      3020      3030      3040      3050      3060      3070

1040      1050      1060      1070      1080      1090      1100
TGTGGCGCGCACCTCATCCAAGTCTCTTCTAACGCTAACATATTTGTCTTTACCTTTTTTAAATCTTTTTT
| | | | | | | | | | | | | | | | | | | | | |
GGTGGTAGTAGAGGATTAGGTGTCTATTATAAAACCAAAA-----CTCATT-CGTTTAATGAA-CTTGACT
3080      3090      3100      3110      3120      3130      3140

1110      1120      1130      1140      1150      1160      1170
TAAATTTAAATTTTATGTGTGTGAGTGTTCCTGCCT-GTATGCACACGTGTGTGTGTGTGTGTGTGA
| | | | | | | | | | | | | | | | | | | | | |
GTCAT---ACCTCTAT-----TTAGT-AAITGCGAGGGTAAGATTCATA-GTAGGAATATTGAAATTTTGG
3150      3160      3170      3180      3190      3200

1180      1190      1200      1210      1220      1230      1240
CACTCTGATCCCTGAGGAGCTCAGAAACAGAAAGGCTTCC--TTCATATAACAACTCCACTTATCCATCCCTC

```

[illegible]



```

| | | | | | | | | | | | | | | | | | | | | |
ATTTAAAGTTTATGAAACATGAAGTGCAG-CTGCAGGATTCTGGCATTCTTGCATGCCATTCTCCATCAGATC
3960      3970      3980      3990      4000      4010      4020

1990      2000      2010      2020      2030      2040      2050      2060
AGATAAAACAACCAAAAGCCTTGACTCCGGTACTAATTCTCCCTGCCGGCCCCGTAAGCATAACGCGGCGA
| | | | | | | | | | | | | | | | | | | | | |
TGGGATGATGGCTCAGAACATGTACACAG--ACTAAGAGTAACTG-TGTGATCTGT-----TAAGGGGTGGA
4030      4040      4050      4060      4070      4080      4090

      2070      2080      2090      2100      2110      2120
TCTCCACTTTAAGAACCT--GGCCGC-GTTCTGCCTGGTCTCGCTTTCTGTAACGGTTCTTACAAAAGTAAT
| | | | | | | | | | | | | | | | | | | | | |
T-AACATAATATGCAGCTTAGGATGTATTTTGTAGATGTAT-GAT-----ATCAGTTCATTC--ACCTGAT
      4100      4110      4120      4130      4140      4150

2130      2140      2150      2160      2170      2180      2190
TAGTTCCTTGCTTTAGCCTCCAAGCTTCTGCTAGTCTATGGCAG-CATCAAGGCTGGT-ATT-TGCTACGGC
| | | | | | | | | | | | | | | | | | | | | |
TACT--TTGGTTGCAGC--ACAA-CTGTATATATTGTATAACCGAAATTGATTATTTTCATTGTCCTTATGC
      4160      4170      4180      4190      4200      4210      4220

      2200      2210      2220      2230      2240      2250      2260
--TGACCGCTA-----CGCCGCCGAATAAGGGTACTGGGCGGCCCGTGAAGGCCCTTTGGTTTCAGAAAC
| | | | | | | | | | | | | | | | | | | | | |
AGTGATTTATAATTAGAGCATGTTTAATAAGTTTACTATTCTTGTAACTA--GTCATTTGACTGGAAAAAA
      4230      4240      4250      4260      4270      4280      4290

      2270      2280      2290      2300      2310      2320      2330
CCAAGGCCCCCTCATACCAACGTTTCGACTTTGATTCTTGCCGGTACGTGGTGGTGGGTGCCTTAGCTCTT
| | | | | | | | | | | | | | | | | | | | | |
ATAAAATACTTTTAAATGGAAAAAAAAAAAAAAAAAAAA
      4300      4310      4320      X

      2340      2350
TCTCGATAGTTAGAC

```

10. ELLIS-012-FIG2AB.SEQ (1-2350)

Q25975 MH mutant porcine ryanodine receptor cDNA.

```

ID Q25975 standard; DNA; 15377 BP.
AC Q25975;
DT 08-JAN-1993 (first entry)
DE MH mutant porcine ryanodine receptor cDNA.
KW MH; RYR1; calcium release channel; sarcoplasmic reticulum;
KW transverse tubule; Pietrain; Yorkshire; polymorphism; beta strand; ss.
OS Synthetic.
FH Key Location/Qualifiers
FT CDS 130..15237
FT /*tag= a
FT variation 207
FT /*tag= b
FT /label= Polymorphic_site
FT variation 405
FT /*tag= c
FT /label= Polymorphic_site
FT variation 438
FT /*tag= d
FT /label= Polymorphic_site
FT variation 876
FT /*tag= e
FT /label= Polymorphic_site
FT variation 1329
FT /*tag= f
FT /label= Polymorphic_site

```

FT variation 1972  
 FT /\*tag= g  
 FT /label= MH\_mutation  
 FT variation 2007  
 FT /\*tag= h  
 FT /label= Polymorphic\_site  
 FT variation 4071  
 FT /\*tag= i  
 FT /label= Polymorphic\_site  
 FT variation 4383  
 FT /\*tag= j  
 FT /label= Polymorphic\_site  
 FT variation 4462  
 FT /\*tag= k  
 FT /label= Polymorphic\_site  
 FT variation 4494  
 FT /\*tag= l  
 FT /label= Polymorphic\_site  
 FT variation 6867  
 FT /\*tag= m  
 FT /label= Polymorphic\_site  
 FT variation 7692  
 FT /\*tag= n  
 FT /label= Polymorphic\_site  
 FT variation 8940  
 FT /\*tag= o  
 FT /label= Polymorphic\_site  
 FT variation 9192  
 FT /\*tag= p  
 FT /label= Polymorphic\_site  
 FT variation 9585  
 FT /\*tag= q  
 FT /label= Polymorphic\_site  
 FT variation 9600  
 FT /\*tag= r  
 FT /label= Polymorphic\_site  
 FT variation 9951  
 FT /\*tag= s  
 FT /label= Polymorphic\_site  
 FT variation 10111  
 FT /\*tag= t  
 FT /label= Polymorphic\_site  
 FT variation 11250  
 FT /\*tag= u  
 FT /label= Polymorphic\_site  
 FT variation 12300  
 FT /\*tag= v  
 FT /label= Polymorphic\_site  
 FT variation 14007  
 FT /\*tag= w  
 FT /label= Polymorphic\_site  
 FT polyA\_signal 15355..15360  
 FT /\*tag= x

PN W09211387-A.

PD 09-JUL-1992.

PF 20-DEC-1991; CA0457.

PR 21-DEC-1990; GB-027869.

PR 20-MAY-1991; GB-010865.

PR 09-SEP-1991; GB-019250.

PA (UYGU-) UNIV GUELPH.

PA (UTOR ) UNIV TORONTO INNOVATIONS FOUND.

PI MacLennan DH, O'Brien PJ;

DR WPI; 92-250106/30.

DR P-PSDB; R25450.

PT Purified DNA mol. for diagnosis of porcine malignant hyperthermia

PT - comprises DNA sequence encoding normal or mutant piggyback

PT receptor with specified endonuclease restriction map

PS Disclosure; Fig 2; 96pp; English.

CC The sequence given is the mutant pig ryanodine receptor (RYR1) gene  
 CC from swine cDNA. The polymorphic sites were observed in comparisons  
 CC of Pietrain and Yorkshire breeds. There are 17 polymorphisms between  
 CC the two breeds. The polymorphism at position 1972 causes a mutation  
 CC from Arg to Cys and this is thought to be the molecular basis of  
 CC porcine malignant hyperthermia (MH). This mutation lies within the  
 CC region of RYR1 that is concerned with the binding of regulators of Ca<sup>2+</sup>  
 CC release channel gating. Analysis of surrounding sequences suggests  
 CC that this mutation lies within a beta strand domain comprising roughly  
 CC of amino acids 520 to 830. RYR1 is the calcium release channel of the  
 CC sarcoplasmic reticulum and is a large protein which spans the gap  
 CC between the transverse tubule and the sarcoplasmic reticulum. The  
 CC channel is activated by ATP, calcium, caffeine, and micro-molar  
 CC ryanodine. It is inhibited by ruthenium red, tetracaine, calmodulin,  
 CC high Mg<sup>2+</sup> and ryanodine.

SQ Sequence 15377 BP; 3197 A; 4630 C; 4755 G; 2774 T;  
 SQ 21 Others;

Initial Score = 134 Optimized Score = 994 Significance = 6.94  
 Residue Identity = 47% Matches = 1215 Mismatches = 988  
 Gaps = 320 Conservative Substitutions = 0

```

                                X      10      20
                                ATGTCCATG--AACTGCTGAGT
                                || | || || || || ||
TTCGAGTAGGGGATGACCTCATCCTCGTCAGTGTCTCCTCTGAGCGTTACCTGCACCTGTGACAGC-CAGT
620      630      640      650      660      670      680

      30      40      50      60      70      80      90
GGATAAACAGCAGC-GGATATCTCTGTCTAAAGGA-ATATTACTACACCAGGAAAAGGACACATTCGACAAC
|| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
GG-GGAGCTCCAGGTTGACGCTCCTTC--ATGCAGACACT-GTGGAAATG--AACCCCATCTGCTCTGGC
690      700      710      720      730      740      750

      100     110     120     130     140     150
AGGAAAGGAGCC--TGTCACAGAAAACACAGTGTCTGTGCATGTGACATTTGCCATGG--GA-AAC--A
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
TGTGAAGAAGGCTATGTGACTGGGGGTGAC---GTCCTCGCCTCTTTACGGACACATGGATGAGTGCCTG
760      770      780      790      800      810      820

      160     170     180     190     200     210     220
ACTGT-TACAACG-TGGTGGTCATTGTGCTGCTGCTAGTGGGCTGTGAGAAGGTGGGAGCCGTGCAGAACTC
|| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
ACCATCTCCCCCGTGACAGTGA-TGACCAGCGCAGACTTGTCTACTACGAGGGKGGATCTGTG-TGCACCC
830      840      850      860      870      880      890

      230     240     250     260     270     280     290
CTG--TGATAACTGT-CAGCCTGGTACTTTCTGCAGAAAATACAATCCAGTCTGCAAGAGCTGCCCT---CC
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
ACGCCCGCTCCCTCTGGAGACTGGAA----CCGCTGAGAATCAGCTGGAGTGGGAGCCACCTGCGCTGGGGC
900      910      920      930      940      950      960

      300     310     320     330     340     350
AAGTACCTTC---TCCAGCATAGGTGGACAGCCGAAGTGAACATCTGCAGAGTG-TGTGCAGG--CTATTT
|| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
CAGCCGCTTCGCATCCGGCAT--GT-CACCACCGGGAGGTACCTGGCGCTCATCGAGGACCGGGCCTGGTG
970      980      990     1000     1010     1020

      360     370     380     390     400     410     420
CAGGTTCAAG-AAGTTTTGCTCCTCTACCCA---CAACGCGGAGTGTGAGTGCATT--GAAGGATTCCATTG
|| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
GTGGTTGATGCCAGCAAGGC-CCAC-ACCAAGGCCACCTCCTTCTGTTCCGCATTTCCAAGGAGAAGCTGG
1030     1040     1050     1060     1070     1080     1090

```

1110            1120            1130            1140            1150            1160            1170  
TTTAAATTTTAT-GTG GTGAGTGTTTGGCTGCCTGTATGC-ACACGTG-TGT-GTGTGTGTGTGTGTGAC  
|| | ||| | ||| | ||| | ||| | ||| ||| ||| ||| |||  
-GCAAGCTGGATCGACTG-GAGGCCT---CCT-CAGGGATCCTGGAGGTGCTGTACTGTGT-CCTGATTGAG  
1780            1790            1800            1810            1820            1830

1180 1190 1200 1210 1220 1230 1240  
ACTCCTGATG-CCTGAGGAGGTCA--GAAGAGAA-AGGGTTGGTTCATAAGAACTGGAGTTATGGA-TGGC  
| | | | | | | | | | | | | | | | | | | | | |  
AGTCCTGAGGTCCTGA--ACATCATCCAGGAGAACCACATCAAGTCCAT--CATCT-CCCTTCTGGACAAGC  
1840 1850 1860 1870 1880 1890 1900

1250 1260 1270 1280 1290 1300  
-TGTGAGCCGGNNGATAGGT-CGGGA---CGGAG--ACCTGTCTTCTTATTTTAACGTGACTGTATAATA  
| | | | | | | | | | | | | | | | | | | | | |  
ATGGGAG--GAACCACAAGGTGCTGGATGTCTGTGTTCCCTGTGTGTGCAATGGTGTGGCCGTGYGCTC  
1910 1920 1930 1940 1950 1960 1970

1310 1320 1330 1340 1350 1360 1370  
AAAAAAAATGATATTTTC-GGGAA-TTGAGAGATTGCTGACACCCCTTCT--AGTTAA---TGATCTAAG  
| | | | | | | | | | | | | | | | | | | | | |  
CAACCAAGATCTCATTACTGAGAACTTGCTG-CCTGGCCGCGAGCTTCTGCTGCAGACAAACCTCATCAACT  
1980 1990 2000 2010 2020 2030 2040

1380 1390 1400 1410 1420 1430 1440  
AGGAATTGTTGATACGTAGTATAGTGTATATGTGTATGTATATGTATATATAAGACTCTTTTACTG  
| | | | | | | | | | | | | | | | | | | | | |  
ATGTCACCAGCATCCGCCCA-AC---ATCTTTGTGGGCCGA-GCAGAGG---GCACCACAC---AGTACAG  
2050 2060 2070 2080 2090 2100

1450 1460 1470 1480 1490 1500 1510  
TCAAAGTCAACCTAGAGTGTC-TGGT-TACCAGGTCAATTTTATT--GGACATTTTACGTACACACACACA  
| | | | | | | | | | | | | | | | | | | | | |  
-CAAATGGTACTTTGAG-GTCATGGTGGACGAAGT-GGTTCCATTCTGACAGCTCAGGCCACCCACCTGCG  
2110 2120 2130 2140 2150 2160 2170

1520 1530 1540 1550 1560 1570  
CACACAC---ACACACAC--ACGTTTATACTACGTA--CTGTTATCG--GTATTCTACGTATATAATGGGA  
| | | | | | | | | | | | | | | | | | | | | |  
GGTGGGCTGGGCCCTACCGAAGGCTACAGCCCCTACCCTGGGGGCGGCGAGGGCTGGGGC-GGCAACGGGG  
2180 2190 2200 2210 2220 2230 2240

1580 1590 1600 1610 1620 1630 1640  
TAGGGTAAAA---GGAAACCAAAG--AGTGA-GTGATATTATTGTGGAGGTGACAGACTACCCCTTCTGGGT  
| | | | | | | | | | | | | | | | | | | | | |  
TCGGCGATGACCTCTATTCTACGGCTTTGACGGGCTGCATCTCTGGACAGGACA-CGTGCCACGCCTGGTG  
2250 2260 2270 2280 2290 2300 2310

1650 1660 1670 1680 1690 1700  
ACGT----AGGGACAGACCTCCTTCGGACTGTCTAAAACTCCCTTAGAAGT-CTCGTCAAGTTCCCGGA-C  
| | | | | | | | | | | | | | | | | | | | | |  
ACTTCCCAGGG-CAG--CACCTTCTGGC-----CCCCGAGGACGTGGTCAGCTGCTGCCTGGACC  
2320 2330 2340 2350 2360 2370

1710 1720 1730 1740 1750 1760 1770  
GAAGAGGACAGAGGAGACAGTCCG-AAAAGTTATTTTCCG-GCAAATCCTTTCCCTGTTTCGTGACACT  
| | | | | | | | | | | | | | | | | | | | | |  
TCAGCGTGCCGTCCA-TCTCCTTCGCGATCAACGGGTGCCCGGTGCAGGGCGTCTTCGAG-GCCTTCAACCT  
2380 2390 2400 2410 2420 2430 2440

1780 1790 1800 1810 1820 1830 1840  
CCAC---CCCTTGTGACACTTGAGTGTATCCT-TGCGCCGGAAGGTC-AGGT--GGTACCCGTCTGTAGG  
| | | | | | | | | | | | | | | | | | | | | |  
CAACGGGCTCTTCTCCCCGT----CGTCAGCTTCTCGGCCGG--TGTCAGGTGCGGTTCC--TCCTTGGG  
2450 2460 2470 2480 2490 2500 2510

1850 1860 1870 1880 1890  
GGCGG---GGAGA---CAGAG-----CCGCGGGGAGCTACGAGAATCGACTCACAGG--GCGCCCCGGG  
| | | | | | | | | | | | | | | | | | | | | |  
GGCGGCCACGGCGAATTCA-AGTTCTCCTCCGCTGGCTACGCCCTTGCCAC-GAGGCTGTGCTC-----  
2520 2530 2540 2550 2560 2570



PD 05-NOV-1991.  
 PF 24-JUN-1988; 212270.  
 PR 24-JUN-1987; US-066078.  
 PR 24-JUN-1988; US-212270.  
 PA (WHIT-) WHITEHEAD INST BIOM.  
 PI Fink GR, Trueheart J, Elion EA;  
 DR WPI; 91-346534/47.  
 DR P-PSDE; R14910.  
 PT DNA fragment contg. pheromone-inducible yeast promoter - useful  
 PT for transforming yeast cells to produce foreign proteins, which  
 PT may be toxic to yeast cells.  
 PS Disclosure; Fig 5; 23pp; English.  
 CC Transcription of the FUS2 gene is greatly enhanced by the presence  
 CC of the appropriate mating pheromone. The promoter region can  
 CC therefore be used for the pheromone inducible expression of proteins  
 CC of interest.  
 CC See also 014754.  
 SQ Sequence 2492 BP; 911 A; 408 C; 441 G; 732 T;

Initial Score = 129 Optimized Score = 939 Significance = 6.63  
 Residue Identity = 46% Matches = 1136 Mismatches = 1038  
 Gaps = 282 Conservative Substitutions = 0

```

                                X      10
                                ATGT---CCATGAACTGCTGAG
                                ||| ||| ||| |||
CTATTGTGCCCGCCGCGTCACAAATGCGCCCGAACTTGTGCGGAAGTTAATCTGAAACAT-ATATGTTACC
130      140      150      160      170      180      190

20      30      40      50      60      70      80
TGGATAAACAGCAGCGGATATCTCTGTCTAAAG-GAA----TATTACTACA-CCAGGAAAAGGACACATTGG
| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
TACTGAAACAGCGCATGTTGGAAAAGACAAAGGTGAAGACGAAGTTGTATATTTAAGATA--GACCCCTTTAT
200      210      220      230      240      250      260

90      100      110      120      130      140      150
ACAAC---AGGAAAGGAGCCTGTACAGAAAACC-ACAGTGTCTGTGCATGTGACATTTGCCATGGGAAA
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
ACATCCTTTTGAAAAAATTATTAATGTGGCAACCGTCTTTTATTTGACAAAGTATCTTTTTCTTTTGAAAA
270      280      290      300      310      320      330      340

160      170      180      190      200      210      220
CAACTGTTACAACGTGGTGGTCATTCT--GCTGCTGCTAGTGGGCTGTGAGAAGGTGGGAGCCGTGCAGAAC
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
CCAATTTTA-GGTTTTCTTTATAGTAAGTTCTTAAGAAAAAGACAAGA-AAACCCCTTGCGATGTTTAAG
350      360      370      380      390      400      410

230      240      250      260      270      280      290
TCCTGTGATAAC-TGTCAGCCTGGTACTTTCTGCAGAA-AATACAATCCAGTCTGCAAGAG-CTGCCCTCCA
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
ACTTCATATAACTTGTACGATTTGAAGTATCCGAAAAATGATTCATTAACGCCAATAAGAGACT--ACAAAA
420      430      440      450      460      470      480

300      310      320      330      340      350      360
AGTAC-CTTCTCCAGCATAGGTGGACAGCCGAAGTGAACATCTGCAGA--GTGTGTGCAGGCTATTTCAGG
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
ATGACTATTTTCATAAAATGATGACAAATTACCAGAAATTGTTAGAAAACCTACGAGAAAGTTAT---CGA
490      500      510      520      530      540      550

370      380      390      400      410      420      430
TTCAAGAAGTTTTGCTCCTCTACCCACAACGC-GGAGTGTGAGTGCATTGAAGGATTCCATTGCTTGGGGCC
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
AACATGAAAACAACTCAACGATAAAAAATTACGAATAAACGACCA-GCAAGTCTGGACTTGCAT---TCT
560      570      580      590      600      610

```

ACAGTGCA---CCAGATGTGAAAAGGACTGCAGGCCCTGGCCAGGAGCTAACGAAGCAGGGTTGCAAAACCTG  
| | | | | | | | | | | | | | | | | | | | | | | | | |  
ATAGTGGAGAGCCTGAGCAATAAAAAATTTA---CTCTCC---TATTAACACAGAGATATTTCAAAA--TG  
620            630            640            650                          660            670            680

510            520            530            540            550            560            570  
TAGCTTGGGA--ACATTTAATGACCAGAACGGTACTGGCGTCTGTGCGACCCTGGACG-AACTGCTCTCTAGA  
| | | | | | | | | | | | | | | | | | | | | | | | | |  
T--CGTGAGACTGAATTTGA-GCCCTCA--GATTCCAATTCTCTCACGAGGGATGCAAATTTTATAAAAT  
690            700            710            720            730            740

580            590            600            610            620            630  
CGGA-AG--GTCTGTGCT-TAAGACCGGGACCACGGAGAAGGACGTGGT--GTGTGGACCCCCTGTGGTGAG  
| | | | | | | | | | | | | | | | | | | | | | | | | |  
CGTACAGGAGTTTTACCTCTCTGAAGTGGGAATA-TTACAATAATTTGTTAACGCAATAACGTATACAGAA  
750            760            770            780            790            800            810            820

640            650            660            670            680            690  
CTTC-TCTCCAGT--ACCACCATTTCTG--TGA-CTCCAGAGGGAGGA--CCAG-GA-GGGC-ACTCCTTG  
| | | | | | | | | | | | | | | | | | | | | | | | | |  
AGGCATTGAATAGTGATCCAAGATTCAAGAATAAACTTGTCAAGCTTGATTCAAGTGACGAGCTATTGCTTT  
830            840            850            860            870            880            890

700            710            720            730            740            750            760  
CAGGTCCTTACCTTGTTCCTGGCGCTGACATCGGCTTTGCTGCTGG-CCCTGATCTTCATTACTCTCCTGTT  
| | | | | | | | | | | | | | | | | | | | | | | | | |  
TTGG---GAACATTGACACTATTGCGTCAATCAGC-AAAATACTGGTAACGGCAATAAAAGAC-CTACGGTT  
900            910            920            930            940            950

770            780            790            800            810            820            830            840  
CTCTGTGCTCAAATGGATCAGGAAAAAATCCCCACATATTCAAGCAACCATTTAAGAAGACCACTGGAGC  
| | | | | | | | | | | | | | | | | | | | | | | | | |  
AGCCAAGC---AACG---TGGGAAAAATGTTGGATGCCA-ATGAATGGCAAAAGATA--TTGACCAAAAATGA  
960            970            980            990            1000            1010            1020

850            860            870            880            890            900            910  
AGCTCAAGAGGAAGATGCTTGTAGCTGCCGATGTCCACAGGAAGAAGAAGGAGGAGGAGGAGGCTATGAGCT  
| | | | | | | | | | | | | | | | | | | | | | | | | |  
GGTACAACA-GCAGCTATATTCAACTTTTGATAT-TTCAG--AGGCGTTCGAGCAACA-TTTGTTAAGA-AT  
1030            1040            1050            1060            1070            1080

920            930            940            950            960            970            980  
GTGATGTA-CTATCCTAGGAGATGTGTGGG---CCGAAACCGAGAAGCACTAGGACCCACCATCCTGTGGA  
| | | | | | | | | | | | | | | | | | | | | | | | | |  
CAAATCCACCTACACAAGCTATTTTGTAGCCACCAAAACAAATGGAAGTA-TTTACTACATTAAGGATGA  
1090            1100            1110            1120            1130            1140            1150

990            1000            1010            1020            1030            1040  
ACAGCACAAGCAACCCACCAC-CCTGTTCT---TACACATCATCCTAGA-TG-ATGTGT---GGGCGCGCA  
| | | | | | | | | | | | | | | | | | | | | | | | | |  
ATA--AGAATCATTTTTTAACAAGTGGTATGAATATTGTTAAAAAGAGAGTGATGTATAAAGTTAGAGGA  
1160            1170            1180            1190            1200            1210            1220

1050            1060            1070            1080            1090            1100            1110  
CCTCAT-CCAAGTCTCTTCTAACG-CTAACATATTTGTCTTTACCTTTTTTAAATCTTTTTTAAAT-TTA-  
| | | | | | | | | | | | | | | | | | | | | | | | | |  
CATATTGAAAAGCCCGATGAAAAGACTGAC-TCAGTGGATTGATACTTTGGAAA-CTTTGGAAGCTGTTAC  
1230            1240            1250            1260            1270            1280            1290

1120            1130            1140            1150            1160            1170  
-AATTTTATGTGT-GTGAGTGTGTTTGCCTGCCTGTATGC---ACACG-TGTGTGTGTGTGTGTGTGTGACAC  
| | | | | | | | | | | | | | | | | | | | | | | | | |  
GAAGATATTCTTTGCCAGAAATTGGGCTTGAAACTAAGCCCGACAAGAAGAAAATAT-TCTTTATTTTCCAA  
1300            1310            1320            1330            1340            1350            1360            1370

1180            1190            1200            1210            1220            1230            1240



T---CCTGATGCCCTGAGGACGT CAGAAGAAAGGGTTGGTTC CAT--AAGAACTGGAGTTATGGATGGCTGT  
| | | | | | | | | | | | | | | | | |  
TAAGTTAGAAACC-GAGG-TCTCCG-AGTATAAGAGT-AATTCCATGTATAATTTCA GTT---TAACCCCAT  
1380 1390 1400 1410 1420 1430

1250 1260 1270 1280 1290 1300 1310  
GAGCCCGNNNGATAGGTCGGGACGGAGACCTGTCTTCTTA--TTTTAACGTGACTGTATAATAAAAAAAAAA  
| | | | | | | | | | | | | | | | | |  
CAGAGATTATACAAAGTTATGATGAAGATCAGTTTACACACCTTTAAAACCCCCAGACAACAAAAATAAA  
1440 1450 1460 1470 1480 1490 1500

1320 1330 1340 1350 1360 1370 1380  
TGATATTTTCGGGAAT-TGTA---GAGA-TTGTCTGCACACCCTTCTAGTTAATGATCT--AAGAGGAATTGT  
| | | | | | | | | | | | | | | | | |  
ATATATGTAATGCATCTCGACAAGAGAGTAATTTGGATAATAGTAGAGTTCCTTCTCTTTCTTTCTGGATCAT  
1510 1520 1530 1540 1550 1560 1570

1390 1400 1410 1420 1430 1440  
TGA--TACGTAGTATACTGTATATGTGTATGTATATGTATATGTATATATAAGACT-CTTTTACTGTCAAAG  
| | | | | | | | | | | | | | | | | |  
CGAGTTAC-TACTCAGATGTATCAGGGCTAGAAATTGT-CACTAATACTTCA-ACTGCCTCAGCTGAGATGA  
1580 1590 1600 1610 1620 1630 1640

1450 1460 1470 1480 1490 1500 1510  
TCAACCT-AGAGTGTCTG-GTTACCAGGTCAATTTTATTGGACATT-TTACGTCACA-CA-CACACACACAC  
| | | | | | | | | | | | | | | | | |  
TAAATCTAAAAATGGATGAAGAAACAG---AATTTTTT---ACATTGGCAGATCACATCAGTAAATTTCAAGA  
1650 1660 1670 1680 1690 1700 1710

1520 1530 1540 1550 1560 1570 1580  
ACACACACACACGTTTATACTACGTACTGTTATCGGTATTCTACGTCATATAATGGGATAGGGTAAAAGGAA  
| | | | | | | | | | | | | | | | | |  
AAGTAATGAAAGGTTTGT--TA-GAA-TTAAAAAAGAATTTATTGAAAAACGATCTGTGCAGGCATTATTGAT  
1720 1730 1740 1750 1760 1770 1780

1590 1600 1610 1620 1630 1640  
ACCAAAGAGTGAGTGATATTATTG--TGG----AGGTGACAGACTAC-----CCCTTCT-GGGTA-CGTAGG  
| | | | | | | | | | | | | | | | | |  
ATC--AGTTAAGAAGAATAAATGCATGGAAAAAGGTGATCGAGTGCGAACGCCCTTCTCGTGCAATTTTTTG  
1790 1800 1810 1820 1830 1840 1850

1650 1660 1670 1680 1690 1700 1710  
GACAGACCTCCTTCGGACTGTCTA-AAACTCCCCTTAGAAGTCTCGTCAAGTTCCCGGACGAAGAGGACAGA  
| | | | | | | | | | | | | | | | | |  
CGCAGATAACTT--AATATCGACCATGTGTTCTTCGTACATAGATAAACTGCATGAACAAAAAATCA-A  
1860 1870 1880 1890 1900 1910 1920

1720 1730 1740 1750 1760 1770 1780  
GGAGACACAGTCCGAAAAGTTATTTTC-CG--GCAAATCCTTT---CCCTGTTTCGTGCACTCCACCCCT  
| | | | | | | | | | | | | | | | | |  
GTA-ACA-ATTTTG-AAACTCACAGAGCTCGAAACAGATGTGATGAACCCACTTGAAAGAATCATAGCCCAT  
1930 1940 1950 1960 1970 1980

1790 1800 1810 1820 1830 1840 1850  
TGT-GGACACTTGAGTGTATCCT---TGCGCCGGGAAGGTCAGGTGGTACCCGTCTGTAGGGCGGGGAGAC  
| | | | | | | | | | | | | | | | | |  
TGTACTACCGTTAAAAG-CAAACATAAAGATTTGCAAGCTTACATGTTA----TTTTTA----CAAGAAAAA  
1990 2000 2010 2020 2030 2040 2050

1860 1870 1880 1890 1900 1910 1920  
AGAGC-CGCGGGGGAGCTACGAGAATCGACTCACAGGCGCGCCCGGCTTCGCAAATGAAACTTTTTTAATC  
| | | | | | | | | | | | | | | | | |  
AAAGCAAATGTGCGAGATATTAACGTGACTTGTGGG---AATGCATTTTCA----AAAC--CTGCAA--  
2060 2070 2080 2090 2100 2110

1930 1940 1950 1960 1970 1980 1990



Sequence 1363 BP; 279 A; 403 C; 368 G; 313 T;

```
Initial Score      =    127  Optimized Score  =    587  Significance =    6.51
Residue Identity  =    48%  Matches          =    709  Mismatches   =    570
Gaps              =    180  Conservative Substitutions =    0
```

310 320 330 340 350 X 360 370  
TCCAGCATAGGTGGACAGCCGAAGTGAACATCTGCAGAGTGTGTGCAGGCTATTTCAAGTTCAAGAAGTTT  
||| | | |||  
GTATAAAGGCCCCCTGGAGCTG  
X 10 20

380            390            400            410            420            430            440  
 TGCTCCTCTACCCACAACGCGGAGTGTGAGTGCA-TTGA--AGGATTCCATTGCTTGGGGCCAC--AGTGCA  
 |||||            |||||            |||||            |||||            |||||            |||||            |||||  
 GGC-CCT---GGCAGAGCCAGAGCTGCAGCCAGATCACCCAGAATCCATTGGCT-----CAACTTTCTGC-  
                   30            40            50            60            70            80

[illegible]

520      530      540      550      560      570      580  
AACATTTAATGACCAGAACGGTACTGGCG-TCTGTGACCCCTGGACGAAGTCTCTCTAGACGGAAGGCTCTG  
|        |        ||     |        |||| |||| |        |||| |||| |        ||||  
AGGCCGT--TCCCCTGGTCACTGTCAAGGCTCTG-CCACCAT-----CCCTGC-CTCTGCCCCGAAGCTTC

160          170          180          190                  200          210

590           600           610           620           630           640  
 TGCTTAAG-ACCGGGACCACG-GA-GAAGGACGTGGTGTGTGGACCCCTG-TGGT-GAGC--TTCTCTCCC  
 ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 TGCTTAAGTCCCTGGAGCAAGTGAGGAAGATCCAGGCCAGCGG-CTCGGTGCTGCTGGAGCAGTTGTGTGCC  
        220       230           240           250           260           270           280

650            660            670            680            690            700  
AGTACCACCATTTCTGTGACTCC-AGAGGGAGGACCAGGAGGG-CACTCCTTG CAGGTC-----CTTACC  
| | | | | | | | | | | | | | | | | | | | | | | | | | | |  
A--CCTACAAGCTGTGTACCCCGAGGAGCTGGTGTGTGCTGGGCCACTCTCTGGGGATCCGAAGGCTTCCC  
290          300          310          320          330          340          350

710            720            730            740            750            760            770  
T-TGTTCTCGGGCT-GACATCGGCTTTGCTGTGGCCCTGATCTTCATTACTCTCCTGTTCTCTGTGCTCA  
| | | | | | | | | | | | | | | | | | | | | | | | | | | |  
TGAGTGGCT-GCTCTAGCCA--GGCCCTGCAGCAGACACAG---TGCCTAAGCCAGCTCCACAGTGGGCTC-  
360            370            380            390            400            410            420

780            790            800            810            820            830            840            850  
AATGGATCAGGAAAAAATTCCCCACATATTCAAGCAACCATTTAAGAAGACCACTGGAGCAGCTCAAGAGG  
|| || | || || || || || || || || || || || || || || || || || || || || || || || || || ||  
--TGCCTC--TACCAAGGTCTCTCGCAGGCTCTATCGGGTATTTCCCTG--CCCTGG---CCCCACCTTG  
               430            440            450            460            470            480

860            870            880            890            900            910  
 AAGATGCTTGTAGCTGCCGATGT--CC-ACA--GGAAGAAGAAGGAGG-AGGAGGAGGCTATGAGCT-GTGA  
 |        |        |        |        |        |        |        |        |        |        |  
 GACTTGCTT-CAGCTG--GATGTTGCCAACTTTGCCACCACCATCTGGCAGCAGATGG--AAAACCTAGGGG  
 490            500            510            520            530            540            550

[illegible]



[illegible]

020 070 040 050 060 070 000

1360            1370            1380            1390            1400            1410            1420

890            900            910            920            930            940            950  
AGAAGGAGGAGGAGGAGGCTATGAGCTGTGATGTACTATCCTAGGAGATGTGTGGGCCGAACCGAGAAGCA  
|| ||||||| || | || | | | | | | | | | | | | | |  
CATATGGGGAGGAAGACGGCAT-CCTTCACAATTCT-TCAGAGAAAACCATTTTGAC-TATGGGTG-AGTA  
1430          1440          1450          1460          1470          1480

960 970 980 990 1000 1010 1020  
CTAGGACCCACCATCCTGTGGAACAGCACAAGCAACCCACCACCCTGTTCTTACATCATCTAGATGA  
|| | | ||| | | | | | | | | | | | | | |  
TGGAGATGTGCTTTGTTGTGCAGGGTCGCTAG--TGGCGTTGACTTGGCCCAGAC-CGTCATCCTTGA-GC  
1490 1500 1510 1520 1530 1540 1550

1030            1040            1050            1060            1070            1080            1090  
TGTGTTGGCGCGCACCTCA-TCCAAGTCTCTTCTAAC-GCT-AACATATTTGTCTTTACCTTTTTTAA--AT  
| |                 | | | | | | | | | | | | | | | | | |  
T-TG-----ACAAGACAGTGGAACAC-CTTCCAACGGCTTGGCAGGTCCATAGGGA-CTGGTTCAATGAT  
                1560            1570            1580            1590            1600            1610

1100      1110      1120      1130      1140      1150      1160  
 CT-TTTTTTAAATTTAAATTTA-TGTGTGTGAGTGTTTTGCCTGCCTGTATGCACACGTGTGTGTGTGT  
 ||    ||    ||    ||    ||    ||    ||    ||    ||    ||    ||    ||    ||    ||    ||    ||  
 CTGGCTCTGCCATGGAACATGAGGGAGCGCAAACTGGAACAACGCAGAAAG-ACTGGT-TGAATTG-GG  
 1620      1630      1640      1650      1660      1670      1680

1170            1180            1190            1200            1210            1220            1230  
GTGTGTGACAC--TCCTGATG--CCTG-AGGAGGTC--AGAAGAGAAAGGGTTGGTCCATAAGAACTGGAG  
| | | | | | | | | | | | | | | | | | | | | |  
GCTCCTCACGCTGTCAAGATGGACGTETACAACCTCGGAGACCAGACTGGAGT-GTTACTGAAG-GCTCTCG  
1690            1700            1710            1720            1730            1740            1750

1240      1250      1260      1270      1280      1290      1300  
 TTATGGATGGCTGTGAGCCGGNNNGATAGGTGGGACGGAGACCTGTCTTCTTA-TTTTAACGTGA-CTGTA  
 | | | ||||| |        | | | | | | | |        | |        | | ||||| |  
 CT-GGGGTTCTGTG-GC---ACACAT-TGAGGGAACCAAGTACCACCTGAAGAGTGCCACGTGACCTGCG  
 1760      1770                  1780      1790      1800      1810      1820

1310      1320      1330      1340      1350      1360  
TAATAAAA---AAAAATG-ATATTTGGGAATT--GTAGAGATTGCTGACACCCTTCTAGTTAATGAT  
| | | | | | | | | | | | | | | | | | | | | |  
AAGTGGGACTGGAAAACTGAAGATGAAAGGTCTTACGTACACAATGT-GTGACA-AAACAAAGTTCA-CAT  
1830      1840      1850      1860      1870      1880      1890

1370 1380 1390 1400 1410 1420 1430  
CTAAGAGGAATTGTTGATAC-GT---AGTATACTGTATATGTGTATGT-ATATGTATATGTATATATAAGA  
||||| | |||| | |||| | |||| | |||| | |||| | |||| | ||||  
GGAAGAGAGCTCCAACAGACAGTGGGCATGATACAGTGGTCATGGAAGTCACAT-TCTCTGGA-ACA-AAG-  
1900 1910 1920 1930 1940 1950 1960

[illegible]

1500            1510            1520            1530            1540            1550            1560  
CGTCACACACACACACACACACACACGT---TT-ATA-CTACGTA-CTGTTATC--GGTATTCTAC  
|| || || || || |         | || |         | || |         | || |         | || |         ||  
CGCCAAACCCA-ACAATTGAAAACA-ATGGAGGTGGCTTCATAGAGATGCAGCTGCCCCCAGGGGAT--AAC  
2030            2040            2050            2060            2070            2080            2090

1570	1560	1550	1400	1410	1420
------	------	------	------	------	------

GTCA--TATAATGGG--ATACGGTAA--AAGCAAACCAAAGAGTGAGTGATATTATTGTGGAGG---TGACA  
 ||| ||| |||| | | || | || | |||| || ||| | | | | ||| | |  
 ATCATCTATGTTGGGGAAGTGAATCAATGGTTCCAAAAGGGAG---CAGCATCG-GAAGGGTTTTCCA  
 2100 2110 2120 2130 2140 2150 2160  
  
 1630 1640 1650 1660 1670 1680 1690  
 GACTACCCCTTCTGGGTACGTAGGGACAGACCTCCTTCGGACTGTCTAAACTCCCCTTAGAAGTCTCGTCA  
 | ||| || | || || ||| || || | | | | ||| || || ||  
 AAAGACCAAGAAAGG---CATA--GAAAGA-CTGACAGTGATAG--GAGAGCACGCCTGGGA---CT--TC-  
 2170 2180 2190 2200 2210 2220  
  
 1700 1710 1720 1730 1740 1750 1760  
 AGTTCCCGGACGAAGAGGACAGAGGACACAGTCCGAAAAGTTATTTTTCCGGCAAATCCTT--TCCCTGT  
 ||| || | || | ||| || | || | | | ||| |||| | | |  
 GGT-CTGCTGGAGGCTTTCTGAG----TTCAATTGGGAAGGCGGTACATACGG----TCCTTGGTGGCGCT  
 2230 2240 2250 2260 2270 2280  
  
 1770 1780 1790 1800 1810 1820 1830  
 TTCGTGACACTCCACCCCTTGTGGACACTTGAGTGT--CA--TCCTTGCGCCGGAAGGTCAGGTGGTACCCG  
 ||| ||| || | ||| | || || | || || || || || || || ||  
 TTC--AACA--GCATCTTCGGGGGAGTGGGGTTTCTACCAAACTTTTATTAGGAGTGGCA-TTGG--CTTG  
 2290 2300 2310 2320 2330 2340  
  
 1840 1850 1860 1870 1880 1890 1900 X  
 TCTGTAGGGGCGGGGA-GACAGAGCCGCGGGGAGCTACGAGAATCGACTCACAGGCGCCCCGGGCTTCGC  
 || || | | || || || | || || || | || || || || || || ||  
 GTTG--GGCCTGAACATGAGAAACCCTACAATG-TCCATGAGCTTTCTCTTGGCTGGAGGTCTGGTCTTCCC  
 2350 2360 2370 2380 2390 2400 2410 X  
  
 1910 1920 1930 1940 1950  
 AAATGAACTTTTTTAATCTCACAAGTTTCGTCCGGGCTCGGCGGACCTA

14. ELLIS-012-FIG2AB.SEQ (1-2350)

035297 ZYMV genome.

ID 035297 standard; DNA; 9593 BP.  
 AC 035297;  
 DT 28-MAY-1993 (first entry)  
 DE ZYMV genome.  
 KW Zucchini yellow mosaic virus; ZYMV; potyvirus; polyprotein; protease;  
 KW proteolytic activity; 49 kD protease; trypsin-like cysteine protease;  
 KW animal picornavirus; sissile bond; Nib; protein; coat; ss.  
 OS Zucchini yellow mosaic virus.  
 FH Key Location/Qualifiers  
 FT 5'UTR 1..139  
 FT /\*tag= a  
 FT CDS 140..9382  
 FT /\*tag= b  
 FT misc\_feature 2437..2438  
 FT /\*tag= c  
 FT /note= "Cleavage site between aphid transmission  
 FT helper component (HC) and the 46 kD protein"  
 FT misc\_feature 3631..3632  
 FT /\*tag= d  
 FT /note= "Cleavage site between 46 kD protein and the  
 FT cytoplasmic inclusion protein (CI)"  
 FT misc\_feature 5533..5534  
 FT /\*tag= e  
 FT /note= "Cleavage site between CI and VPg/protease (VPg  
 FT and protease are probably not separated in  
 FT ZYMV)"  
 FT misc\_feature 6991..6992  
 FT /\*tag= f  
 FT /note= "Cleavage site between VPg/protease and RNA  
 FT replicase (REP)"



FT misc\_feature 8542..8543  
 FT /\*tag= g  
 FT /note= "Cleavage site between REP and the coat  
 FT protein (CP)"  
 FT misc\_feature 9382  
 FT /\*tag= h  
 FT /note= "Polyprotein termination point"  
 FT 3'UTR 9383..9593  
 FT /\*tag= i  
 PN W09301305-A.  
 PD 21-JAN-1993.  
 PF 09-JUL-1992; U05745.  
 PR 09-JUL-1991; US-727837.  
 PA (BALI/) BALINT R.  
 PI Balint R;  
 DR WPI; 93-045506/05.  
 DR P-PSDB; R35081.  
 PT Method for identifying protease inhibitors - useful for drugs  
 PT screening for treating e.g. chronic inflammation, metastatic  
 PT cancers and viral infections  
 PS Disclosure; Fig 4; 62pp; English.  
 CC This sequence represents the nucleotide sequence of the zucchini yellow  
 CC mosaic virus (ZYMV) genome. ZYMV is a potyvirus and expresses its  
 CC genome as a single 350 kD polyprotein which is cleaved into at least  
 CC seven mature gene products by three distinct proteolytic activities.  
 CC Two of the proteases are virus encoded, including the potyviral 49 kD  
 CC protease. This protease is responsible for at least five of the seven  
 CC cleavages. This enzyme is a trypsin-like cysteine protease which is  
 CC structurally and mechanistically representative of the largest class  
 CC of viral proteases, including those of the animal picornaviruses.  
 CC This enzyme is highly specific and appears to recognise a region  
 CC comprised of about seven amino acids surrounding the scissile bond. Of  
 CC the five sites cleaved by this enzyme, the two flanking the protease  
 CC appear to be cleaved intramolecularly, while the remaining three  
 CC appear to be cleaved intermolecularly. Of the latter three, the site  
 CC between the NIb protein and the coat protein appears to be the most  
 CC active. The polyprotein sequence encoded by this genome is not  
 CC given in the specification but is deduced in R35081.  
 SQ Sequence 9593 BP; 2995 A; 1844 C; 2258 G; 2496 T;

Initial Score = 124 Optimized Score = 977 Significance = 6.32  
 Residue Identity = 47% Matches = 1213 Mismatches = 969  
 Gaps = 353 Conservative Substitutions = 0

```

                                X      10      20
                                AT--GTCCATGAACTGCTGAGT
                                || ||| ||| ||| |||
GTCAAGGATTTATTCACTTCTGGTGTGAAACACAGAGCAAGCGAGAAAGATGGGTCTACGAA-AGCTGTGA
6960      6970      6980      6990      7000      7010      7020

      30      40      50      60      70      80
GGATAAACAGCACGGGATAT---CTCTGTCTAA--AGGAATA-TTACTACACCAG--GAAAAGGACACAT-T
| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
AGGGAACC--TTCGGGCTGTTGGAAGTGCACAATCAGCGTTAGTCACCAAAACATGTTGTGAAAGGCAAGTGT
7030      7040      7050      7060      7070      7080      7090

      90      100     110     120     130     140     150
CGACAACAGGAAAGGAGCCTGTCACAGAAAACACAGTGTCTGTGCATGTGACATTTGGC-CATGGGAAAC
| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
CCTTTCTTCGAA--GAATAT-TTACAACACACGCAGAAGCGAGCGCCTATTTAGACCCCTAATGGGAGAG
7100      7110     7120     7130     7140     7150     7160

      160     170     180     190     200     210     220
AACTGTTACAACGTGGTGGTCATTG-TGCTGCTGCTAGTGGG---CTGTGAGAAGGTGGGAGCCGTGCAGA
|| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
TACCAGCCGACCAAGTTGAACAAAGAACCC---CTTTAAAAACCATTTCTTTAAATACAATAAACCCGT-CA--

```

7170      7180      7190                  7200      7210      7220      7230

230                  240                  250                  260                  270                  280  
ACTCCTGTGA--TAACTGTCAGCCTGGT-AC TTTCTG---CAGAAAAT---ACAATCC--AGTCTGCAAGA  
| | | | | | | | | | | | | | | | | | | | | | | |  
----CTGTTAACCAACTG-GATCATGATAAATTTTTGGGAGCAGTGGATGGGTTATACGTATGATGTGTGA  
7240                  7250                  7260                  7270                  7280                  7290

290                  300                  310                  320                  330                  340  
GCTGCCCTCCAAGTACCTTC--TCCAGCATAGGTGGACAGCGGAACGT--AAC-ATCTGCAGAGTGTGTGC  
| | | | | | | | | | | | | | | | | | | | | | | |  
TTTTGAGTTCAACGAATGTCGATTACATTACAGAT--CCCGAGGAAATTTACAACCTCTTTGAACA-TGAAAGC  
7300                  7310                  7320                  7330                  7340                  7350                  7360

350                  360                  370                  380                  390                  400                  410                  420  
AGGCTATTTTCAGGTTCAAGAAGTTTTGCTCCTCTACCCACAACGCGGAGTGTGAGTGCATTGAAGGATTCCA  
| | | | | | | | | | | | | | | | | | | | | | | |  
A-GCAATTGGA-GCCCA----GTATAG---AGGAAAGAAGAAAGAGTATTTTGAGGGGCTAG-ATGATT---  
7370                  7380                  7390                  7400                  7410                  7420

430                  440                  450                  460                  470                  480  
TTGCTTG-GGGCCAC----AGTGCACCAGATGTGAAAAAGGACTGC-AGGCCTGGCCAGGAGCTAACGAA-GC  
| | | | | | | | | | | | | | | | | | | | | | | |  
TTGATCGAGAGCGACTTTTATTCCA-AAGTTGTGAAAGGTTGTTCAATGGCT-ACAAAGGTCTGTGGAATGG  
7430                  7440                  7450                  7460                  7470                  7480                  7490

490                  500                  510                  520                  530                  540  
AGGGTT----GCAAAAC-CTG-TAGCTTGGGAACATTTAATG--ACCAGAAC-GGTAC--TGGCGTCTGTGC  
| | | | | | | | | | | | | | | | | | | | | | | |  
ATCTTTAAGGCCGAGCTCAGGCCGCTTGAGAA-AGTCAGGCTAACAAAACACGAACCTTTACAGCAG-CG  
7500                  7510                  7520                  7530                  7540                  7550                  7560

550                  560                  570                  580                  590                  600                  610  
ACCCTGGACGAAGTCTCTCTAGACGGAAGGTCTGTGCTTAAGACCGGGACCACGGAGAAGGACGTGGTGTG  
| | | | | | | | | | | | | | | | | | | | | | | |  
CCAATTGATACATTGCT---TGGAGCTAAAGTTTGTGTGGATGATTTCAACAATG-----AGTTC-TACAGGA  
7570                  7580                  7590                  7600                  7610                  7620                  7630

620                  630                  640                  650                  660                  670                  680  
TGGACCCCCTGTGGTGAGCTTCTCTC-CCAGTACCACCATTTCT-GTGA CTCCAGAGGGAGGA--CCAGGAG  
| | | | | | | | | | | | | | | | | | | | | | | |  
AAAACCTCAAGTGTCCATGGACGGTCGGCATGACAAAATTTATGGTGGTT-----GGGATAAATTGATGAG  
7640                  7650                  7660                  7670                  7680                  7690

690                  700                  710                  720                  730                  740                  750  
GGCACTCCTTGCAGG-TCCTTACCTTGTTCCTGGCGCTGACATCGGCTTTGCTGCTGGCCCTGATCTTCATT  
| | | | | | | | | | | | | | | | | | | | | | | |  
ATCATTACCTGATGGTTGGTTGTATTG-TCATGCTGATG-GATC-ACAGTTCGATAGTTCTGTTAACCCCA-G  
7700                  7710                  7720                  7730                  7740                  7750                  7760

760                  770                  780                  790                  800                  810  
ACTCTCCTGTTCTCTGTGCTCA-AAT--GG-----ATCAGGAAAA-----AATTCCCCC-ACATATTCAA  
| | | | | | | | | | | | | | | | | | | | | | | |  
CCT-TACTGAACGCAGTGTCTATAATCAGGTCA TTTTATATGGAGGATTGGTGGGTCGGCCAAGAGATGCTT  
7770                  7780                  7790                  7800                  7810                  7820                  7830

820                  830                  840                  850                  860                  870                  880  
GCAACCATTTAAGAAGACCACTG-GAGCAGCTCAAGAGGAAGATGCTTGTAGCTGCCGATGTCCACAGGAA-  
| | | | | | | | | | | | | | | | | | | | | | | |  
GAAAATCTTTATGCCGA-GATTGTGTACA-CTCCAATTCTTGCTCCTGATGG--AACAATTTTCA-AGAAAT  
7840                  7850                  7860                  7870                  7880                  7890                  7900

890                  900                  910                  920                  930                  940                  950  
GAAGAAGGAGGAGGAGGAGGCTA---TGAGCTGTGATGTACTATCCTAGGAGATGTGTGGGCCGAAACCGAG  
| | | | | | | | | | | | | | | | | | | | | | | |  
TTAGACCTAACAACAGTCCCAACGCTCAACAGTCTGCA-TAACACACTA-ATCGTCTC-----ATCTGC

7910            7920            7930            7940            7950            7960  
               960                970                980                990                1000                1010  
 AAGCACTAGGACCCCA--CCATCCTGT-GGAACAGC-ACAAGCAACCCACCACCCTG--TTC-TTACACAT  
 |    | | | |    |    | |    | | | | | | | |    | |    | | | | | | | | | |  
 ATTTACTATGCGTGCATGAAATTTGGTTGGAAGTGCAGGAGATTGAGAATAAACTTGTCTTCTTTGCAAAT  
 7970            7980            7990            8000            8010            8020            8030            8040

8650      8660      8670      8680      8690      8700      8710  
1690      1700      1710      1720      1730      1740      1750  
GTCTCGTCAAGTTCCCGGACGAAGAGGACAGAGGAGACACAGTCCGAAAAGTTATTT--TTC-CGGCA--AA  
| | | | | | | | | | | | | | | | | | | | | | | | | | | |  
CTTTTCG--AAGATCAC-AAAGAAAATGTCA-TTGCCACGC-GT--GAAAGGAAATGTGATACTCGATATTGA  
8720      8730      8740      8750      8760      8770  
1760      1770      1780      1790      1800      1810  
TCCTTTCCCTGTTTCGTGACACTCCACCCCTTGTG-----GACACTTGAGTGTATCCTTGCGCCGGAAGG  
| | | | | | | | | | | | | | | | | | | | | | | | | | | |  
TCATTT-GCTGGAATATAAACCGGATCAAATTGAGTTATATAACACACGAGCGTC-TCAT--CAGCAGTTTCG  
8780      8790      8800      8810      8820      8830      8840  
1820      1830      1840      1850      1860      1870      1880  
TCAGGTGGT---ACC-CGTCTGTAGGGCGGGGA---GACAGAGCCGCGG--GGGAGCTACGAGAATCGACT  
| | | | | | | | | | | | | | | | | | | | | | | | | | | |  
CCTCTTGGTTCAACCAGGTTAAGACGGAATATGATTTGAACGAGCAACAGATGGGAGTTGTAATGAATG-GT  
8850      8860      8870      8880      8890      8900      8910  
1890      1900      1910      1920      1930      1940      1950  
CACAGGGCGCCCCGGGCTTCGCAATGAAACTTTTTTAATCTCACA--AGTTTCGTCCGGGCT-CGGCGGAC  
| | | | | | | | | | | | | | | | | | | | | | | | | | | |  
TTCATG--GTTTGGTCATTGAGAATGGCAC----TTCACCGGACATTAATGGAGTGTGGTTATGATGGAC  
8920      8930      8940      8950      8960      8970      8980  
1960      1970      1980      1990      2000      2010      2020  
CTATGGCGTCGATCCTTATTACCTTATCC--TGGCGCCAAGATAAAACAACCAAAAGCCTTGACTCCGG-TA  
| | | | | | | | | | | | | | | | | | | | | | | | | | | |  
---GGAAATGAGC--AAGTTGAGTATCCCTTGAAACCAATAGTTGAAATGCAAAGCCAACGCTGCGGCAA  
8990      9000      9010      9020      9030      9040  
2030      2040      2050      2060      2070      2080  
CTAATTC-TC--CCTGCCG-GCCCCCGTAAGCATACGCGGCGATCTCCACTTTAAGAACCTGGCCGCGTTC  
| | | | | | | | | | | | | | | | | | | | | | | | | | | |  
ATAATGCATCATTTTTTTCAGATGCAGCGGAGGCAT-ATATAGAGAT-GAGAAATGCAGA----GGCACCATAC  
9050      9060      9070      9080      9090      9100      9110  
2090      2100      2110      2120      2130      2140      2150  
-TGCC-TGGTCTCGCTTTTCGTAACGGTTCTTACAAAAGTAATTAGTTCTTGCTTTTCAGC-CTCCAAGCTTC  
| | | | | | | | | | | | | | | | | | | | | | | | | | | |  
ATGCCGAGGTATGTTTTCCTT---CGAAACCTAC----GGGAT-AG-----GAGTTTACGATATGCTTT  
9120      9130      9140      9150      9160      9170  
2160      2170      2180      2190      2200      2210      2220  
TGCTAGTCTATGGCAGCA--TC-AAGGCTGGTATTTGCTACGGCTGA-CCGCTACGCCGCCGCAATAAG-GG  
| | | | | | | | | | | | | | | | | | | | | | | | | | | |  
TGAT-TTCTATGAAGTCAATTCTAAACTCCTGAAAGAGCCCGCGAAGCTGTTGCGCAGATGAAAGCAGCAG  
9180      9190      9200      9210      9220      9230      9240  
2230      2240      2250      2260      2270      2280      2290  
TACTGGGC---GGCCCGTCGAAGGCCCTTTGG-TTTCA--GAAACCAAGGCCCCCTCATACCAACGTTTC  
| | | | | | | | | | | | | | | | | | | | | | | | | | | |  
CTCTTAGCAATGTTTCTTC-AAGGTTGTTTGGCCTTGATGGAAA--TGTTGCCACC-----ACTAGCG--AA  
9250      9260      9270      9280      9290      9300  
2300      2310      2320      2330      2340      2350  
GACTTTGATTCTTGC-CGGTACGTGGTG---GTGGGTGCCT-TAGCTCTTTCTCGATAGTTAGAC  
| | | | | | | | | | | | | | | | | | | | | | | | | | | |  
GACACTGAACGGCACACTGCACGTGATGTTAATAGAAACATGCACACCTTACTAGGTGTGAATACAATGCAG  
9310      9320      9330      9340      9350      9360      9370 X  
TAAAGGGTAGGCCGCCTACCTAGGTTATTGTTTCGCTGCCGAC  
9380      9390      9400      9410      9420

[illegible]

ACCCA---ACCCT-CCAC--AAACCCATGATTTTTCTTGGCTAATATGTCATTTCTGGAGATTGGTATG  
 160 170 180 190 200 210 220

160 170 180 190 200 210 220  
 CAACTGTTACAA--CGT--GGTGGTCATTGTGCTGCTAGTGGGCTGTGAGAAGGTGGGAGCCGTGCAGA  
 ||||| | | | | | | | | | | | | | | | | | |  
 TCACTGTTACGATTCTAAGATGCTCGCTG-GCTTC-ATTGGT-TCCAAGGAGAACCATGGA-CAG--CTGA  
 230 240 250 260 270 280

230 240 250 260 270 280  
 ACTCCTGTG----ATAACTGTCAGCCTGGTACTTTCTGCAGAAAATACAATCCAGTCTGCAAGAGCTGC--C  
 ||||| | | | | | | | | | | | | | | | | | |  
 TCTCCTTTGAGGCATGCATGACACAACCTCTACTTTTTCTGGGCTTGGGTTGCACAGAGTGTCTCTTTG  
 290 300 310 320 330 340 350 360

290 300 310 320 330 340 350  
 CTCCAAGTACCTTCTCCAGCATAGGTGGACAGCCG-----AACTGTA--ACATCTGCAGAGTGTG-GCAGG  
 || | | | | | | | | | | | | | | | | | | | | |  
 CTGTGATGGCCTATGACCGC-TATGTGGCTATCTGTCATCCACTCCACTACCCCGTCATTGTCAGTAGCCGG  
 370 380 390 400 410 420 430

360 370 380 390 400 410 420  
 CTATTTCAAGTTCAAGAAGTTTGTCTCTTAC-CCACAACGCGGAGTGTGAGTGCATTGAAGGATTCC--A  
 |||| | | | | | | | | | | | | | | | | | | |  
 CTATGTGTGCAGATGGCAG-CTGGATCCTGGGCTGGAGGTTTTGGTATCTCCATG-GTTAAAGTTTTCTTA  
 440 450 460 470 480 490 500

430 440 450 460 470 480  
 TTGCTTG--GGGC-CACAGTGCACCAGATGTGAAAAGGACTGCAGGCTGGCCAGGAGCTAACGAAGCAGGG  
 || | | | | | | | | | | | | | | | | | | | | |  
 TTTCTGCCTGTCTTACTGTGGCCCCAACACCATCA--AC--CA--CTTTTCTG----TGATGTGTCTCCA  
 510 520 530 540 550 560

490 500 510 520 530 540 550  
 TTGCAAAACCTGTAGCTTGGGAACATTTA-ATGAC--CAGAACGG--TACTGGCGTCTGTGACCCCTGGACG  
 |||| | | | | | | | | | | | | | | | | | | |  
 TTGCTCAACCTGT--CATG----CACTGACATGTCCACAGCACAGCTTACAGAC-TTGTG----CCTGG-CC  
 570 580 590 600 610 620

560 570 580 590 600 610 620  
 AACTGCTCTTAGACGGA-AGGTCTGTGCTTAAGACCGGACCACGGAGAAGGACGT----GGTG-TGTGGA  
 | | | | | | | | | | | | | | | | | | | | | |  
 ATTTTATTCTGCTGGGACCGCTCTCTG-TCA---CTGGGCATCCTACATGGCCATCACAGGTGCTGTGAT  
 630 640 650 660 670 680 690

630 640 650 660 670 680  
 CCCCTGTGGTGAGCTTCTCTCC-CAGTACCACCATTT----CTGTGACTCCAGAGGGAGGACCAG--GAGG  
 | | | | | | | | | | | | | | | | | | | | | |  
 GCGCATCCCCTCAGCTGCTGGCGCCATAAAGCCTTTCAACCTGTGCCTCCACCTCA----CTGTTGTGA  
 700 710 720 730 740 750

690 700 710 720 730 740 750  
 GCA-CTCCT-TGCAG---GTCCTTACCT-TGTTCTGGCGCTGACATCGGCTTTGCTGCTGGCCCTGATCTT  
 || | | | | | | | | | | | | | | | | | | | | |  
 TCATCTTCTATGCAGCCAGTATTTTCATCTATGCCAGGC-CTAAGGCACTCTCAGCTTTAGACACCAACAAG  
 760 770 780 790 800 810 820 830

760 770 780 790 800 810 820  
 CATTACTCTCCTGTTCTGTGCT-CAAATGGATCAGGAAAAAATCCCCACATA-TTCAAG--CAACCAT  
 | | | | | | | | | | | | | | | | | | | | | |  
 C--TGGTCT-CTGACTCTACGCTGTGATGTACCGTTGTTCAATCCCATCATCTACTGCTTGGCGAACCC--  
 840 850 860 870 880 890

830 840 850 860 870 880 890  
 TTAAGAAGACCACTGGAGCAGCTCAAGAGGAAGATGCTTGTAGCT-GCCGATGTCCACAGGAAG--AAGAAG  
 |||| | | | | | | | | | | | | | | | | | | |

--AAGATG-TCAAAAGAGC-GCT-ACGTGG--CACGC-TGCACCTGGCCCAGGAC--CAGGAGGCCAATACC  
900 910 920 930 940 950  
900 910 920 930 940 950 960  
GAGGAGGAGGAGGCTATGAGCTGTGATGTACTATCCTAGGAGATGTGTGGGCCGAAACCGAGAAGCACTAGG  
| | ||| || | |||  
AACAA--AGGCAGC-AAAATTGGTTAG  
960 970 980 X

ACCCC

> 0 <  
0| |0 IntelliGenetics  
> 0 <

FastDB - Fast Pairwise Comparison of Sequences  
Release 5.4

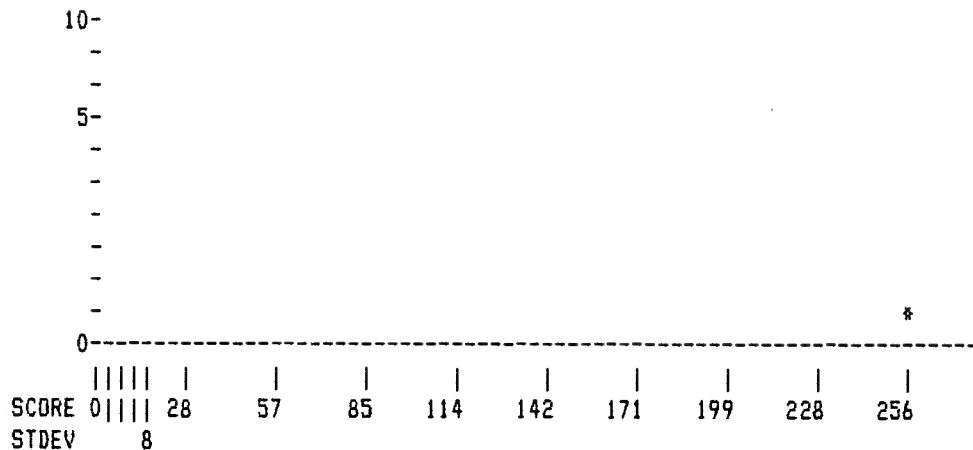
Results file ellis-012-fig2ab-pir.res made by shears on Tue 14 Sep 93 15:01:23-PDT.

Query sequence being compared: ELLIS-012-FIG2AB.PEP (1-256)  
Number of sequences searched: 52257  
Number of scores above cutoff: 4100

Results of the initial comparison of ELLIS-012-FIG2AB.PEP (1-256) with:  
Data bank : PIR 36, all entries

100000-  
-  
N -  
U50000-  
M -\*  
B -  
E -  
R -  
\*  
D -  
F10000-  
-  
S -  
E 5000-  
Q -  
U -  
E -  
N -  
C -  
E -  
S 1000-  
-  
-  
500-  
-  
-  
-  
-  
- \*  
-  
-  
100-  
-  
-  
50-  
-  
-  
-  
-  
-





# PARAMETERS

Similarity matrix	Unitary	K-tuple	2
Mismatch penalty	5	Joining penalty	30
Gap penalty	1.00	Window size	32
Gap size penalty	0.26		
Cutoff score	0		
Randomization group	0		
Initial scores to save	40	Alignments to save	15
Optimized scores to save	0	Display context	50

# SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	4	5	1.51

Times:	CPU	Total Elapsed
	00:03:05.07	00:06:20.00

Number of residues:	15485766
Number of sequences searched:	52257
Number of scores above cutoff:	4100

Cut-off raised to 4.  
 Cut-off raised to 5.  
 Cut-off raised to 6.  
 Cut-off raised to 7.

The scores below are sorted by initial score.  
 Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was found:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
1. B32393	4-1BB protein precursor - Mou	256	256	256	166.38	0

The list of other best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
**** 4 standard deviations above mean ****						
2. S15785	Heat-stable antigen HSA-C - M	141	11	26	4.62	0
3. A39046	Tissue factor precursor - Mou	294	11	42	4.62	0
4. A32318	Tissue factor precursor - Mou	294	11	42	4.62	0
**** 3 standard deviations above mean ****						
5. S15783	Heat-stable antigen precursor	145	10	15	3.84	0

6. S01877	NADH dehydrogenase (ubiquinon	59	10	14	3.96	0
7. S15784	Heat-stable antigen - Mouse	76	10	19	3.96	0
8. A43537	Heat stable antigen M1/69-J11	76	10	19	3.96	0
9. S21969	19K zein precursor (clone ZG3	214	10	19	3.96	0
10. Z12MA2	19K zein precursor (clone c71	230	10	23	3.96	0
11. S03417	19K zein precursor (clone gZ1	234	10	23	3.96	0
12. S21970	19K zein precursor (clone A30	234	10	23	3.96	0
13. Z12MB1	19K zein precursor (clone cZ1	234	10	23	3.96	0
14. Z12M3	19K zein precursor (clone A30	234	10	23	3.96	0
15. S15655	Zein, 19K - Maize	235	10	23	3.96	0
16. Z12M99	19K zein precursor (clone ZG9	235	10	22	3.96	0
17. S07172	19K zein precursor (clone Z4)	267	10	21	3.96	0
18. BWMSV4	Mov-34 protein - Mouse	321	10	38	3.96	0
19. S27672	O-antigen polymerase - Salmon	359	10	21	3.96	0
20. A32118	H+-transporting ATP synthase	465	10	39	3.96	0
21. S01292	Tenascin - Chicken (fragment)	697	10	36	3.96	0
22. C33379	Protenascin 190K precursor -	1535	10	36	3.96	0
23. B32230	Cytotactin precursor 2 - Chic	1537	10	36	3.96	0
24. B33379	Protenascin 200K precursor -	1626	10	36	3.96	0
25. A30903	Protenascin precursor - Chick	1808	10	36	3.96	0
26. A33379	Protenascin 230K precursor -	1808	10	36	3.96	0
27. A32230	Cytotactin precursor - Chicke	1810	10	36	3.96	0
28. B39079	Pre-alpha-inhibitor HC3 chain	18	9	9	3.30	0
29. C34245	Inter-alpha-trypsin inhibitor	20	9	9	3.30	0
30. B25604	Endothelial cell growth facto	49	9	9	3.30	0
31. D31201	GLI-related finger protein HK	106	9	23	3.30	0
32. S12586	Whey acidic protein - Rabbit	127	9	16	3.30	0
33. S01286	Whey acidic protein precursor	127	9	16	3.30	0
34. S03552	Inter-alpha-trypsin inhibitor	147	9	9	3.30	0
35. B30020	Hypothetical protein 6 - Frui	174	9	15	3.30	0
36. S01189	NADH dehydrogenase (ubiquinon	174	9	16	3.30	0
37. S19934	Hypothetical protein - Escher	196	9	28	3.30	0
38. A42337	submandibular gland protein A	206	9	30	3.30	0
39. A25303	Alpha-1-microglobulin precurs	220	9	34	3.30	0
40. TVNST2	Transforming protein (int-2)	245	9	34	3.30	0

# 1. ELLIS-012-FIG2AB.PEP (1-256)

B32393 4-1BB protein precursor - Mouse

ENTRY B32393 #Type Protein  
TITLE 4-1BB protein precursor - Mouse  
DATE 17-Jul-1992 #Sequence 17-Jul-1992 #Text 23-Mar-1993  
PLACEMENT 0.0 0.0 0.0 0.0 0.0  
SOURCE Mus musculus #Common-name house mouse  
ACCESSION B32393  
REFERENCE  
#Authors Kwon B.S., Weissman S.M.  
#Journal Proc. Natl. Acad. Sci. U.S.A. (1989) 86:1963-1967  
#Title cDNA sequence of two inducible T-cell genes.  
#Reference-number A32393; MUID:89184547  
#Accession B32393  
#Molecule-type mRNA  
#Residues 1-256 <KWD>  
#Cross-reference GB:J04492

FEATURE  
1-23 #Domain signal sequence (predicted)  
<SIG>\  
24-256 #Protein 4-1BB protein <MAT>  
SUMMARY #Molecular-weight 27598 #Length 256 #Checksum 4884  
SEQUENCE

Initial Score = 256 Optimized Score = 256 Significance = 166.38  
Residue Identity = 100% Matches = 256 Mismatches = 0  
Gaps = 0 Conservative Substitutions = 0

```

X      10      20      30      40      50      60      70
MGNNCYNVVVIVLLL VGCEKVGAVQNSDCNCQPGTFCRKYNPVCKSCPPSTFSSIGGQPNICRV CAGYFR
|||||
MGNNCYNVVVIVLLL VGCEKVGAVQNSDCNCQPGTFCRKYNPVCKSCPPSTFSSIGGQPNICRV CAGYFR
X      10      20      30      40      50      60      70

      80      90      100      110      120      130      140
FKKFCSSSTHNAECEIEGFHCLGPQCTRCEKDCRPGQELTKQGCKTCSLGT FNDQNGTGVC RPWTNCSLDGR
|||||
FKKFCSSSTHNAECEIEGFHCLGPQCTRCEKDCRPGQELTKQGCKTCSLGT FNDQNGTGVC RPWTNCSLDGR
      80      90      100      110      120      130      140

      150      160      170      180      190      200      210
SVLKTGTTEKDVVCGPPVVSFSPSTTISVTPGGPGGHSLOVLTFLALTSALLLALIFITLLFSVLKWIRK
|||||
SVLKTGTTEKDVVCGPPVVSFSPSTTISVTPGGPGGHSLOVLTFLALTSALLLALIFITLLFSVLKWIRK
      150      160      170      180      190      200      210

      220      230      240      250      X
KFPHFIFKQPFKKTGAAQEEEDACSCRCPOEEEGGGGYEL
|||||
KFPHFIFKQPFKKTGAAQEEEDACSCRCPOEEEGGGGYEL
      220      230      240      250      X

```

2. ELLIS-012-FIG2AB.PEP (1-256)

S15785 Heat-stable antigen HSA-C - Mouse

```

ENTRY      S15785      #Type Protein
TITLE      Heat-stable antigen HSA-C - Mouse
DATE       07-Apr-1992 #Sequence 07-Apr-1992 #Text 07-Apr-1992
PLACEMENT  0.0 0.0 0.0 0.0 0.0
SOURCE     Mus musculus #Common-name house mouse
ACCESSION  S15785
REFERENCE
#Authors   Wenger R.H., Ayane M., Bose R., Koehler G., Nielsen
            P.J.
#Journal    Eur. J. Immunol. (1991) 21:1039-1046
#Title      The genes for a mouse hematopoietic differentiation
            marker called the heat-stable antigen.
#Reference-number S15783; MUID:91209380
#Accession  S15785
#Status     preliminary
#Residues   1-141 <WEN>
#Cross-reference EMBL:X56486
SUMMARY     #Molecular-weight 15515 #Length 141 #Checksum 6244
SEQUENCE

```

```

Initial Score      =      11  Optimized Score      =      26  Significance      =      4.62
Residue Identity   =      22%  Matches              =      30  Mismatches       =      93
Gaps               =          9  Conservative Substitutions      =          0

```

```

      90      100      110      120      130 X      140      150
NAECECIEGFHCLGPQCTRCEKDCRPGQELTKQGCKTCSLGT FNDQNGTGVC RPWTNCSLDGRSVLKTGTTE
                                     |   |   |
                                     MGRAMVARLGLGLLLALLLPT
                                     X      10      20

      160      170      180      190      200      210
KDVVCGPPVVSFSPSTTISVTP-----EGGPGGHSLOVLTFLAL--TSALLLALIFITLLFSVLKWIRKKF
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QIYCNQTSVAPFSGNQNISASPNPSNATTRGGSSLOSTAGLLALSSTSLLETQARKRLYFPIFYTPKWQ
      30      40      50      60      70      80      90

```

```

220      230      240      250      X
PWTEKQPEKKTGAAQEEEDACSCRCPOEEEGGGGYEL

```

| | | | |  
P--QVQCDQEEETGPPRIVCYHTSTENTENSKFDGIKGRVKGLREERCY  
100 110 120 130 140

### 3. ELLIS-012-FIG2AB.PEP (1-256)

A39046 Tissue factor precursor - Mouse

ENTRY A39046 #Type Protein  
TITLE Tissue factor precursor - Mouse  
DATE 31-Jul-1991 #Sequence 31-Jul-1991 #Text 23-Mar-1993  
PLACEMENT 0.0 0.0 0.0 0.0 0.0  
SOURCE Mus musculus #Common-name house mouse  
ACCESSION A39046  
REFERENCE  
#Authors Ranganathan G., Blatti S.P., Subramaniam M., Fass  
D.N., Maihle N.J., Getz M.J.  
#Journal J. Biol. Chem. (1991) 266:496-501  
#Title Cloning of murine tissue factor and regulation of  
gene expression by transforming growth factor type  
beta1.  
#Reference-number A39046; MUID:91093171  
#Accession A39046  
#Status preliminary  
#Molecule-type mRNA  
#Residues 1-294 <RAN>  
#Cross-reference GB:J05713  
SUMMARY #Molecular-weight 32935 #Length 294 #Checksum 8911  
SEQUENCE

Initial Score = 11 Optimized Score = 42 Significance = 4.62  
Residue Identity = 23% Matches = 58 Mismatches = 165  
Gaps = 28 Conservative Substitutions = 0

X 10 20  
MGNNCYNVVVIVLLLVGCEK-V  
| | | |  
VRPRLLAALAPTFLGCLLLQVIAGAGIPEKAFNLTWISTDFKTILEWQPKPTNYTYTVQISDRSRNWNKNCF  
10 20 30 40 50 X 60 70  
30 40 50 60 70 80 90  
GAVQNSCDNCQPGTFCRKYNPVCK-SCPPSTFSSIGGQPNICRVCAGYFRFKKF-CSSTHNAECECIEGF  
| | | | | | | | | |  
STTDTECDLTDEIVKDVWYAEAKVLSVPRRNSVHGDGDLVIHGEEPPFTNAPKFLPYRDTNLGQPVIOQF  
80 90 100 110 120 130 140  
100 110 120 130 140 150  
HCLGPQCTRCEKDCRPGQELTKQGCKTCSLGTFNQNG--TGVCRPWTNCSLDGRSVLKTGTTE--KDVVCG  
| | | | | | | | | | | |  
EQDGRKLNVVVKD-----SLT-LVRKNGTFLTLRQVFGKDLGYIITYRKGSSTGKKTNITNTNEFSIDVEEG  
150 160 170 180 190 200 210  
160 170 180 190 200 210 220  
PPVVSFSPSTTIS-VTPEGGPGGHSLOVLT----LFLALT-----SALLLALIFITLLFSVLKWKIRKKFPHI  
| | | | | | | | | | | |  
VSYCFFVQAMIFSRKTNQNSPG--SSTVCTEQWKSFLGETLIIVGAVVLLATIFIILLSISLCKRRK---NR  
220 230 240 250 260 270 280  
230 X 240 250  
FKQPFKKTGTGAQEEEDACSCRCPEEEGGGGGYEL  
| | |  
AGQKGKNTPSRLA  
290 X

A32318 Tissue factor precursor - Mouse

ENTRY A32318 #Type Protein  
TITLE Tissue factor precursor - Mouse  
DATE 29-Jan-1990 #Sequence 29-Jan-1990 #Text 23-Mar-1993  
PLACEMENT 0.0 0.0 0.0 0.0 0.0  
SOURCE Mus musculus #Common-name house mouse  
ACCESSION A32318  
REFERENCE  
#Authors Hartzell S., Ryder K., Lanahan A., Lau L.F., Nathans D.  
#Journal Mol. Cell. Biol. (1989) 9:2567-2573  
#Title A growth factor-responsive gene of murine BALB/c 3T3 cells encodes a protein homologous to human tissue factor.  
#Reference-number A32318; MUID:89343974  
#Accession A32318  
#Status preliminary  
#Molecule-type mRNA  
#Residues 1-294 <HAR>  
#Comment This sequence has not been compared to the nucleotide translation.  
SUMMARY #Molecular-weight 32923 #Length 294 #Checksum 9197  
SEQUENCE

Initial Score = 11 Optimized Score = 42 Significance = 4.62  
Residue Identity = 23% Matches = 58 Mismatches = 165  
Gaps = 28 Conservative Substitutions = 0

```

                                     X      10      20
                                     MGNNCYNVVIVLLLVGCEK-V
                                     | | |
VRPRLAALAPTFLGCLLLQVTAGAGIPEKAFNLTHWISTDFKTIIEWQPKPTNYTYTVQISDRSRNWKNKCF
 10      20      30      40      50      X 60      70

      30      40      50      60      70      80      90
GAVQNSCDNCQPGTFCKRKNPVCK-SCPPSTFSSIGGQPNPNCICRVACGYFRFKKF-CSSTHNAECECIEGF
  ||      | | | | | | | | | | | | | | | | | |
STTDTECDLTDEIVKDVWAYEAKVLSVPRRNSVHGDGDLVIHGEEPPFTNAPKFLPYRDTNLGQPVIGQF
 80      90     100     110     120     130     140

      100     110     120     130     140     150
HCLGPQCTRCEKDCRPQGLTKQGCKTCSLGTFNQNG--TGVCRPWTNCSLDGRSVLKTGTTE--KDVVCG
 | | | | | | | | | | | | | | | | | | | | | |
EQDGRKLNVVVKD-----SLT-LVRKNGTFLTLRQVFGKDLGYIITYRKGSSTGKKTNITNTNEFSIDVEEG
150     160     170     180     190     200     210

160     170     180     190     200     210     220
PPVVSFSPSTTIS-VTPEGGPGHSLQVLT----LFLALT-----SALLLALIFITLLFSVLKWIRKKFPHI
 | | | | | | | | | | | | | | | | | | | | | |
VSYCFVQAMIFSRKTQNSPG--SSTVCTEQWKSFLGETLIIVGAVVLLATIFIILLSISLCKRRK---NR
 220     230     240     250     260     270     280

      230      X 240     250
FKQPFKKTGTGAQEEEDACSCRCPEEEGGGGGYEL
 | | |
AGQKGKNTPSRLA
 290      X
```

5. ELLIS-012-FIG2AB.PEP (1-256)

S15783 Heat-stable antigen precursor - Mouse

ENTRY S15783 #Type Protein  
TITLE Heat-stable antigen precursor - Mouse  
DATE 28-Aug-1992 #Sequence 28-Aug-1992 #Text 28-Aug-1992

PLACEMENT 0.0 0.0 0.0 0.0 0.0  
 SOURCE Mus musculus #Common-name house mouse  
 ACCESSION S15783  
 REFERENCE  
 #Authors Wenger R.H., Ayane M., Bose R., Koehler G., Nielsen P.J.  
 #Journal Eur. J. Immunol. (1991) 21:1039-1046  
 #Title The genes for a mouse hematopoietic differentiation marker called the heat-stable antigen.  
 #Reference-number S15783; MUID:91209380  
 #Accession S15783  
 #Status preliminary  
 #Residues 1-45 <WEN>  
 #Cross-reference EMBL:X53825  
 SUMMARY #Molecular-weight 4485 #Length 45 #Checksum 9465  
 SEQUENCE

Initial Score = 10 Optimized Score = 15 Significance = 3.96  
 Residue Identity = 36% Matches = 17 Mismatches = 28  
 Gaps = 2 Conservative Substitutions = 0

```

      110      120      130      140      150      160      170
PGQELTKQGCKTCSLGTFNQNGTGVCRPWTNCSLDGRSVLKTGTTEKDVVCGPPVVSFSPSTTISVTPEGG
                                     |  |  |  |
                                     APFPGNQNISASPNPSNATTRG
                                     X    10    20

180      190      200 X  210      220      230      240      250
PGGHSLSQVLTLFLALTSALLLALIFITLLFSVLKWIRKKFPHIFKQPFKKTGAAGQEDACSCRCPOEEEGG
|| |||  ||| |  || |
-GGSSLQSTAGLLAL-SLSLLHLYC
      30      40      X
  
```

GGG

#### 6. ELLIS-012-FIG2AB.PEP (1-256)

S01877 NADH dehydrogenase (ubiquinone) chain 5 - Brine  
 ENTRY S01877 #Type Protein (fragment)  
 TITLE NADH dehydrogenase (ubiquinone) chain 5 - Brine  
 shrimp mitochondrion (SGC4) (fragment) #EC-number  
 1.6.5.3  
 DATE 31-Mar-1990 #Sequence 31-Mar-1990 #Text 23-Mar-1993  
 PLACEMENT 0.0 0.0 0.0 0.0 0.0  
 SOURCE mitochondrion Artemia sp. #Common-name brine shrimp  
 ACCESSION S01877  
 REFERENCE  
 #Authors Batuecas B., Garesse R., Calleja M., Valverde J.R., Marco R.  
 #Journal Nucleic Acids Res. (1988) 16:6515-6529  
 #Title Genome organization of Artemia mitochondrial DNA.  
 #Reference-number S01207; MUID:88289417  
 #Accession S01877  
 #Molecule-type DNA  
 #Residues 1-59 <BAT>  
 #Cross-reference EMBL:X07663  
 KEYWORDS mitochondrion\ oxidoreductase  
 GENETIC  
 #Special-code 4  
 SUMMARY #Length 59 #Checksum 9192  
 SEQUENCE

Initial Score = 10 Optimized Score = 14 Significance = 3.96  
 Residue Identity = 35% Matches = 15 Mismatches = 42

Gaps = 2 Conservative Substitutions = 0

```

110      120      130      140      150      160      170      180
ELTKQGCKTCSLGTNDQNGTGVCPRWTNCSLDGRSVLKTGTTEKDVVCGPPVVSFSPSTTISVTPEGGPGG
                                         | |
                                         MGELLYHEGDCGHWEEAGPSLI
                                         X   10   20

      190      200      210      X 220      230      240      250
H--SLQVLTFLALTSALLLALIFITLLFSVLKWIRKKFPHIFKQPFKKTGAQEEEDACSCRCPQEEEGGG
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
HNSLRGSSLSFLTSPPYKVLILSSLLFTLFMYSMA
      30      40      50      X

GGYEL

```

# 7. ELLIS-012-FIG2AB.PEP (1-256)

S15784 Heat-stable antigen - Mouse

```

ENTRY      S15784      #Type Protein
TITLE      Heat-stable antigen - Mouse
DATE       07-Apr-1992 #Sequence 07-Apr-1992 #Text 07-Apr-1992
PLACEMENT  0.0 0.0 0.0 0.0 0.0
SOURCE     Mus musculus #Common-name house mouse
ACCESSION  S15784
REFERENCE
#Authors   Wenger R.H., Ayane M., Bose R., Koehler G., Nielsen
            P.J.
#Journal    Eur. J. Immunol. (1991) 21:1039-1046
#Title      The genes for a mouse hematopoietic differentiation
            marker called the heat-stable antigen.
#Reference-number S15783; MUID:91209380
#Accession  S15784
#Status     preliminary
#Residues   1-76 <WEN>
#Cross-reference EMBL:X56469
SUMMARY     #Molecular-weight 7797 #Length 76 #Checksum 2479
SEQUENCE

```

Initial Score = 10 Optimized Score = 19 Significance = 3.96  
 Residue Identity = 28% Matches = 22 Mismatches = 52  
 Gaps = 4 Conservative Substitutions = 0

```

      80      90      100      110      120      130      140
SSTHNAECEEIEGFHCLGPQCTRCEKDCRPGQELTKQGCKTCSLGTNDQNGTGVCPRWTNCSLDGRSVLKT
                                         | | |
                                         MGRAMVARLGLGLLLALLPT
                                         X   10   20

      150      160      170      180      190      200 X 210
--GTTEKDVVCGPPVVSFSPSTTISVTPEGGPGGHSQVLTFLALTSALLLALIFITLLFSVLKWIRKKFP
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QIYCNQTSVAPFPGNQISASPNPSNATTRG-GGSSLQSTAGLLAL-SLSLLHLYC
      30      40      50      60      70      X

      220      230      240      250
HIFKQPFKKTGAQEEEDACSCRCPQEEEGGGG

```

# 8. ELLIS-012-FIG2AB.PEP (1-256)

A43537 Heat stable antigen M1/69-J11d precursor - Mouse

```

ENTRY      A43537      #Type Protein
TITLE      Heat stable antigen M1/69-J11d precursor - Mouse

```

DATE 06-Nov-1992 #Sequence 06-Nov-1992 #Text 23-Mar-1993  
 PLACEMENT 0.0 0.0 0.0 0.0 0.0  
 SOURCE Mus musculus #Common-name house mouse  
 ACCESSION A43537  
 REFERENCE  
 #Authors Kay R., Takei F., Humphries R.K.  
 #Journal J. Immunol. (1990) 145:1952-1959  
 #Title Expression cloning of a cDNA encoding M1/69-J11d  
 heat-stable antigens.  
 #Reference-number A43537; MUID:90361906  
 #Accession A43537  
 #Status preliminary  
 #Molecule-type mRNA  
 #Residues 1-76 <KAY>  
 #Cross-reference GB:M58661  
 #Comment This sequence has not been compared to the  
 nucleotide translation.  
 SUMMARY #Molecular-weight 7797 #Length 76 #Checksum 2479  
 SEQUENCE

Initial Score = 10 Optimized Score = 19 Significance = 3.96  
 Residue Identity = 28% Matches = 22 Mismatches = 52  
 Gaps = 4 Conservative Substitutions = 0

80 90 100 110 120 130 140  
 SSTHNAECECIEGFHCLGPQCTRCEKDCRPGQELTKQGCKTCSLGTFTNDQNGTGVCRPWTNCSLDGRSVLKT  
 | | | |  
 MGRAMVARLGLGLLLALLPT  
 X 10 20

150 160 170 180 190 200 X 210  
 --GTTEKDVVCGPPVVSFSPSTTISVTPEGGPGHSLQVLTLFLALTSALLLALIFITLLFSVLKWIRKKFP  
 | | | | | | | | | | | | | | | |  
 QIYCNQTSVAPFPGNQISASPNPSNATTRG-GGSSLQSTAGLLAL-SLSLLHLYC  
 30 40 50 60 70 X

220 230 240 250  
 HIFKQPFKKTGAAGEEDACSCRCPOEEEGGGGG

#### 9. ELLIS-012-FIG2AB.PEP (1-256)

S21969 19K zein precursor (clone ZG31A) - Maize (fragment)

ENTRY S21969 #Type Protein (fragment)  
 TITLE 19K zein precursor (clone ZG31A) - Maize (fragment)  
 DATE 04-Dec-1992 #Sequence 04-Dec-1992 #Text 23-Mar-1993  
 PLACEMENT 0.0 0.0 0.0 0.0 0.0  
 SOURCE Zea mays #Common-name maize  
 ACCESSION S21969  
 REFERENCE  
 #Authors Hu N.T., Peifer M.A., Heidecker G., Messing J.,  
 Rubenstein I.  
 #Journal EMBO J. (1982) 1:1337-1342  
 #Title Primary structure of a genomic zein sequence of  
 maize.  
 #Reference-number S07172; MUID:84207882  
 #Accession S21969  
 #Molecule-type mRNA  
 #Residues 1-214 <HUN>  
 #Cross-reference EMBL:V01473  
 #Comment The translation of the nucleotide sequence is not  
 given in this paper.  
 SUPERFAMILY #Name zein  
 KEYWORDS seed\ storage protein  
 SUMMARY #Length 214 #Checksum 4377  
 SEQUENCE



Initial Score = 10 Optimized Score = 19 Significance = 3.96  
 Residue Identity = 21% Matches = 27 Mismatches = 83  
 Gaps = 16 Conservative Substitutions = 0

```

      90      100      110      120      130 X      140      150
AECECIEGFHCLGPQCTRCEKDCRPGQELTKQGCKTCSLGTENDQNGTGVCRPWTNCSLDGRSVLKTGTTEK
                                     ||      |
                                     ATIFPQCSQAPIASLLPPYLSP
                                     X      10      20

```

```

      160      170      180      190      200      210      220
DV--VCGPPVVSFSPSTTISVTPGEGPGGHSLOVLTFLALTSALLLALIFITLLFSVLKWIRKKFPHIFKQ
| || |      |      |      |      ||| ||| |      ||      ||      |
AVSSVCENP--ILQPYRIQQAITAG-----ILPLSPLFLQSSALLHQLPLVHLL---AQNIR---AQQQLQ
      30      40      50      60      70      80

```

```

      230      240      250      X
PFKKTGAAQEEEDACSCRCPEEEGGGGGYEL
      ||      |      |
QLVLANLAAYSQSQQLFPFNQLAALNSASYLQQQLPFSQLPAAYPQQFLPFNQLAALNSPAYLQQQLLPF
      90      100      110 X      120      130      140      150

```

SQLAGVSPAT  
 160

#### 10. ELLIS-012-FIG2AB.PEP (1-256)

ZIZMA2 19K zein precursor (clone cZ19A2) - Maize (fragment)

ENTRY ZIZMA2 #Type Protein (fragment)  
 TITLE 19K zein precursor (clone cZ19A2) - Maize (fragment)  
 DATE 30-Jun-1988 #Sequence 30-Jun-1988 #Text 31-Mar-1993  
 PLACEMENT 1340.0 1.0 4.0 3.0 1.0  
 SOURCE Zea mays #Common-name maize  
 ACCESSION D24557  
 REFERENCE (Inbred line W64A)  
 #Authors Marks M.D., Lindell J.S., Larkins B.A.  
 #Journal J. Biol. Chem. (1985) 260:16451-16459  
 #Title Nucleotide sequence analysis of zein mRNAs from  
 maize endosperm.  
 #Reference-number A92510; MUID:86059563  
 #Accession D24557  
 #Molecule-type mRNA  
 #Residues 1-230 <MAR>  
 #Comment The authors translated the codon GAC for residue 209  
 as Asn.  
 SUPERFAMILY #Name zein  
 KEYWORDS seed\ storage protein  
 FEATURE  
 1-18 #Domain signal sequence (fragment) <SIG>\  
 19-230 #Protein 19K zein <MAT>  
 SUMMARY #Length 230 #Checksum 8546  
 SEQUENCE

Initial Score = 10 Optimized Score = 23 Significance = 3.96  
 Residue Identity = 22% Matches = 33 Mismatches = 94  
 Gaps = 19 Conservative Substitutions = 0

```

      70      80      90      100      110 X      120      130
ICRVCAGYFRFKKCSSTHNAECECIEGFHCLGPQCTRCEKDCRPGQELTKQGCKTCSLGTENDQNGTGVCR
                                     | |      ||      |
                                     KIFCFLMLLG-LSASAATATIF
                                     X      10      20

```

```

      140      150      160      170      180      190      200

```



LPFNQLAALNSPAYLQQQQLLPFSQLAGVSPAT  
150 160 170 180

## 12. ELLIS-012-FIG2AB.PEP (1-256)

S21970 19K zein precursor (clone A30) - Maize

ENTRY S21970 #Type Protein  
TITLE 19K zein precursor (clone A30) - Maize  
DATE 04-Dec-1992 #Sequence 04-Dec-1992 #Text 23-Mar-1993  
PLACEMENT 0.0 0.0 0.0 0.0 0.0  
SOURCE Zea mays #Common-name maize  
ACCESSION S21970  
REFERENCE  
#Authors Hu N.T., Peifer M.A., Heidecker G., Messing J.,  
Rubenstein I.  
#Journal EMBD J. (1982) 1:1337-1342  
#Title Primary structure of a genomic zein sequence of  
maize.  
#Reference-number S07172; NUID:84207882  
#Accession S21970  
#Molecule-type mRNA  
#Residues 1-234 <HUN>  
#Cross-reference EMBL:V01481  
#Comment The translation of the nucleotide sequence is not  
given in this paper.  
SUPERFAMILY #Name zein  
KEYWORDS seed\ storage protein  
FEATURE  
1-21 #Domain signal sequence <SIG>\  
22-234 #Protein 19K zein <MAT>  
SUMMARY #Molecular-weight 25403 #Length 234 #Checksum 977  
SEQUENCE

Initial Score = 10 Optimized Score = 23 Significance = 3.96  
Residue Identity = 22% Matches = 33 Mismatches = 97  
Gaps = 19 Conservative Substitutions = 0

60 70 80 90 100 110 120 130  
NCNICRVCAGYFRFKKFCSSTHNAEECEIEGFHCLGPQCTRCEKDCRPGQELTKQGCKTCSLGTENDQNGTG  
| | | |  
MAAKIFCLLMLLG-LSASAATA  
X 10 20

140 150 160 170 180 190 200  
VCRPWTNCSLDGRSVLKTGTTEKDV--VCGPPVVSFSPSTTISVTPEGGPGGHSLOVLTFLALTSALLLAL  
| | | | | | | | | | | | | | | |  
TIFF--QCSQAPIASLLPPYLSFAVSSVCENP--ILQPYRIQQAIAAG-----ILPLSPLFLQSSALLQQL  
30 40 50 60 70 80

210 220 230 240 250 X  
IFITLLFSVLKWKIRKFKPHIFKQPFKKTGAAQEEDACSCRCPEEEGGGGGYEL  
| | | | | | | | | |  
PLVHLL---AQNIR----AQQLQQLVLANLAAYSQQQQFLPFNQLAALNSASYLQQQQLPFSQLPAAYPQQF  
90 100 110 120 130 X 140

LPFNQLAALNSPAYLQQQQLLPFSQLAGVSPAT  
150 160 170 180

## 13. ELLIS-012-FIG2AB.PEP (1-256)

Z1ZMB1 19K zein precursor (clone cZ19B1) - Maize

ENTRY Z1ZMB1 #Type Protein

TITLE 19K zein precursor (clone cZ19B1) - Maize  
 DATE 30-Jun-1988 #Sequence 30-Jun-1988 #Text 31-Mar-1993  
 PLACEMENT 1340.0 1.0 4.0 2.0 2.0  
 SOURCE Zea mays #Common-name maize  
 ACCESSION E24557  
 REFERENCE (Inbred line W64A)  
 #Authors Marks M.D., Lindell J.S., Larkins B.A.  
 #Journal J. Biol. Chem. (1985) 260:16451-16459  
 #Title Nucleotide sequence analysis of zein mRNAs from  
 maize endosperm.  
 #Reference-number A92510; MUID:86059563  
 #Accession E24557  
 #Molecule-type mRNA  
 #Residues 1-234 <MAR>  
 SUPERFAMILY #Name zein  
 KEYWORDS seed\ storage protein  
 FEATURE  
 1-21 #Domain signal sequence <SIG>\  
 22-234 #Protein 19K zein <MAT>  
 SUMMARY #Molecular-weight 25435 #Length 234 #Checksum 3129  
 SEQUENCE

Initial Score = 10 Optimized Score = 23 Significance = 3.96  
 Residue Identity = 22% Matches = 33 Mismatches = 97  
 Gaps = 19 Conservative Substitutions = 0

```

60      70      80      90     100     110     120     130
NCNICRVCAGYFRFKKFCSSTHNAEECEICGFHCLGPQCTRCEKDCRPGQELTKQGCKTCSLGTfNDQNGTG
                                     | | |
                                   MAAKIFCLMLLG-LSASAATA
                                   X      10      20

140     150     160     170     180     190     200
VCRPWTNCSLDGRSVLKTGTTEKDV--VCGPPVVSFSPSTTISVTPEGGPGGHSLOVLTFLALTSALLLAL
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
TIFP--QCSQAPIASLLPPYLSSAVSSVCENP--ILQPYRIQQAIAAG-----ILPLSPLFLQSSALLQQL
      30      40      50      60      70      80

210     220     230     240     250     X
IFITLLFSVLKWKIRKKFPHIFKQPFKKTGAAQEEEDACSCRCPQEEEGGGGYEL
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
PLVHLL---AQNIR----AQQLQQLVLANLAAYSQSQQLFPFNQLGSLNSASYLQSQQLPFSQLPAAYPQQF
      90     100     110     120     130 X     140

LPFNQLAALNSPAYLQSQQLLPFSQLAGVSPAT
150     160     170     180
  
```

#### 14. ELLIS-012-FIG2AB.PEP (1-256)

ZIZM3 19K zein precursor (clone A30) - Maize

ENTRY ZIZM3 #Type Protein  
 TITLE 19K zein precursor (clone A30) - Maize  
 DATE 18-Dec-1981 #Sequence 30-Jun-1988 #Text 31-Mar-1993  
 PLACEMENT 1340.0 1.0 4.0 2.0 1.0  
 SOURCE Zea mays #Common-name maize  
 ACCESSION C22762\ A03349  
 REFERENCE (Clone A30, sequence translated from the mRNA  
 sequence)  
 #Authors Geraghty D., Peifer M.A., Rubenstein I., Messing J.  
 #Journal Nucleic Acids Res. (1981) 9:5163-5174  
 #Title The primary structure of a plant protein: zein.  
 #Reference-number A93741; MUID:82081837  
 REFERENCE (Revision to amino end)  
 #Authors Geraghty D.E., Messing J., Rubenstein I.  
 #Journal ENRD J. (1982) 1:1329-1335

```

#Title      Sequence analysis and comparison of cDNAs of the
            zein multigene family.
#Reference-number A90967; MUID:84207881
SUPERFAMILY #Name zein
KEYWORDS    seed\ storage protein
FEATURE
  1-21      #Domain signal sequence <SIG>\
  22-234    #Protein 19K zein <MAT>
SUMMARY     #Molecular-weight 25403 #Length 234 #Checksum 977
SEQUENCE

Initial Score      =    10  Optimized Score =    23  Significance = 3.96
Residue Identity   =   22%  Matches         =    33  Mismatches  =  97
Gaps               =    19  Conservative Substitutions =    0

      60          70          80          90         100         110         120         130
NCNICRVCAGYFRFKKFCSSSTHNAECECIEGFHCLGPQCTRCEKDCRPGQELTKQGCKTCSLGTFNQNGTG
                                     | |  ||  |
                                     MAAKIFCLMLLG-LSASAATA
                                     X    10    20

      140         150         160         170         180         190         200
VCRPWTNCSLDGRSVLKTGTTEKDV--VCGPPVVSFSPSTTISVTPEGGPGGHSLOVLTFLALTSALLLAL
|  ||  |  ||  |  ||  |  ||  |  ||  |  ||  |
TIFP--QCSQAPIASLLPPYLSPAVSSVCENP--ILQPYRIQQAIAAG-----ILPLSPLFLQSSALLQQL
      30          40          50          60          70          80

      210         220         230         240         250         X
IFITLLFSVLKWKIRKKFPHIFKQPFKKTGAAQEEEDACSCRCPEEEEGGGGYEL
||  ||  |  ||  |  ||  |
PLVHLL---AQNIR----AQQQLQLVLANLAAYSQ0000FLPFNQLAALNSASYLQ000LPFSQLPAAYPQQF
      90          100         110         120         130 X    140

LPFNQLAALNSPAYLQ000QLLPFSQLAGVSPAT
150          160          170          180

```

# 15. ELLIS-012-FIG2AB.PEP (1-256)

S15655 Zein, 19K - Maize

```

ENTRY      S15655      #Type Protein
TITLE      Zein, 19K - Maize
DATE       04-Apr-1992 #Sequence 04-Apr-1992 #Text 04-Apr-1992
PLACEMENT  0.0  0.0  0.0  0.0  0.0
SOURCE     Zea mays #Common-name maize
ACCESSION  S15655
REFERENCE
#Authors   Quayle T.J.A., Brown J.W.S., Feix G.
#Journal    Gene (1989) 80:249-257
#Title      Analysis of distal flanking regions of maize 19-kDa
            zein genes.
#Reference-number S15655; MUID:90060774
#Accession  S15655
#Status     preliminary
#Residues   1-235 <QUA>
#Cross-reference EMBL:X53582
SUMMARY     #Molecular-weight 25505 #Length 235 #Checksum 1651
SEQUENCE

```

```

Initial Score      =    10  Optimized Score =    23  Significance = 3.96
Residue Identity   =   22%  Matches         =    33  Mismatches  =  97
Gaps               =    19  Conservative Substitutions =    0

```

```

      60          70          80          90         100         110         120         130
NCNICRVCAGYFRFKKFCSSSTHNAECECIEGFHCLGPQCTRCEKDCRPGQELTKQGCKTCSLGTFNQNGTG

```

MAAKIFCLLMLLG-LSASAATA  
X 10 20

140 150 160 170 180 190 200  
VCRPWTNCSLDGRSVLKTGTTEKDV--VCGPPVVSFSPSTTISVTPEGGPGGHSLSQVLTFLALTSALLLAL  
| | | | | | | | | | | | | | | |  
TIFP--QCSQAPIASLLPPYLSPAVSSVCENP--ILQPYRIQQAIAG-----ILPLSPLFLQQSSALLQQL  
30 40 50 60 70 80

210 220 230 240 250 X  
IFITLLFSVLKWIRKKFPHIFKQPFKKTGAAQEEEDACSCRCPQEEEEGGGGYEL  
| | | | | | | | | | | | | | | |  
PLVHLL---AQNIR---AQQLQQLVLANVAAYSQQQQFLPFNQLAALNSAAYLQQQQLLPFSQLTAAYPQQ  
90 100 110 120 130 X 140

FLPFNQLAALNSAAYLQQQQLLPFSQLAVVSPA  
150 160 170 180

> 0 <  
0| 0D IntelliGenetics  
> 0 <

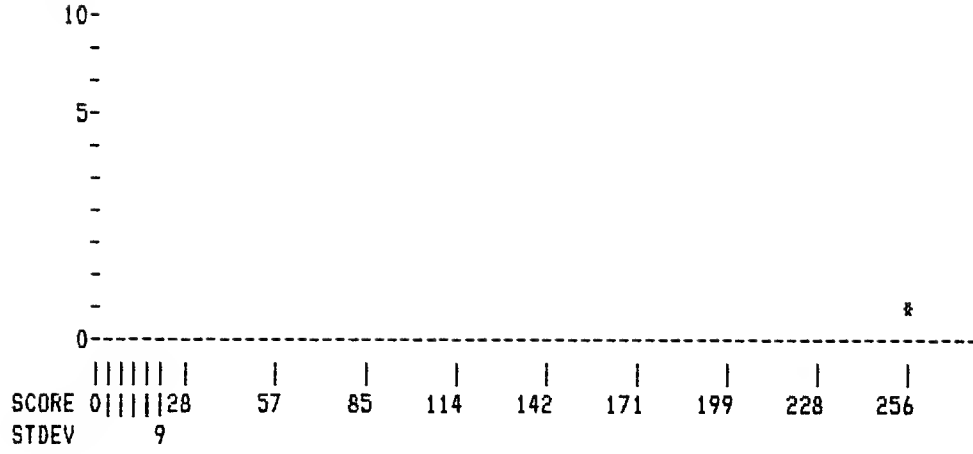
FastDB - Fast Pairwise Comparison of Sequences  
Release 5.4

Results file ellis-012-fig2ab-spt.res made by shears on Tue 14 Sep 93 15:06:00-PDT.

Query sequence being compared: ELLIS-012-FIG2AB.PEP (1-256)  
Number of sequences searched: 29955  
Number of scores above cutoff: 3792

Results of the initial comparison of ELLIS-012-FIG2AB.PEP (1-256) with:  
Data bank : Swiss-Prot 25, all entries

100000-  
-  
N -  
050000-  
M -  
B -  
E -%  
R -  
-  
O -  
F10000-  
\*  
S -  
E 5000-  
Q -  
U -  
E -  
N -  
C -  
E -  
S 1000-  
-  
-  
500-  
-  
-  
-  
-  
- \*  
-  
100-  
-  
-  
50-  
-  
-  
-  
-  
-



PARAMETERS

Similarity matrix	Unitary	K-tuple	2
Mismatch penalty	5	Joining penalty	30
Gap penalty	1.00	Window size	32
Gap size penalty	0.26		
Cutoff score	0		
Randomization group	0		
Initial scores to save	40	Alignments to save	15
Optimized scores to save	0	Display context	50

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	4	5	1.75
Times:	CPU	Total Elapsed	
	00:01:57.02	00:03:57.00	
Number of residues:	10214020		
Number of sequences searched:	29955		
Number of scores above cutoff:	3792		

Cut-off raised to 4.  
Cut-off raised to 5.  
Cut-off raised to 6.  
Cut-off raised to 7.

The scores below are sorted by initial score.  
Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was found:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
1. 41BB_MOUSE	T CELL ANTIGEN 4-1BB PRECURSOR	256	256	256	143.96	0

The list of other best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
**** 3 standard deviations above mean ****						
2. TF_MOUSE	TISSUE FACTOR PRECURSOR (TF)	294	11	42	4.00	0
3. NUSM_ARTSX	NADH-UBIQUINONE OXIDOREDUCTAS	59	10	14	3.43	0
4. M169_MOUSE	M1/69-J11D HEAT STABLE ANTIGEN	76	10	19	3.43	0
5. ZEAB_MAIZE	ZEIN-ALPHA PRECURSOR (19 KD)	186	10	21	3.43	0
6. ZFA3_MAIZE	ZEIN-ALPHA PRECURSOR (19 KD)	270	10	23	3.43	0



7. ZEAS_MAIZE	ZEIN-ALPHA PRECURSOR (19 KD)	234	10	23	3.43	0
8. ZEA4_MAIZE	ZEIN-ALPHA PRECURSOR (19 KD)	234	10	23	3.43	0
9. ZEA1_MAIZE	ZEIN-ALPHA PRECURSOR (19 KD)	234	10	23	3.43	0
10. ZEAC_MAIZE	ZEIN-ALPHA PRECURSOR (19 KD)	235	10	23	3.43	0
11. ZEA2_MAIZE	ZEIN-ALPHA PRECURSOR (19 KD)	235	10	22	3.43	0
12. ZEAL_MAIZE	ZEIN-ALPHA PRECURSOR (CLONE Z	253	10	21	3.43	0
13. MD34_MOUSE	MDV34 PROTEIN.	321	10	38	3.43	0
14. ATPB_SULAC	MEMBRANE-ASSOCIATED ATPASE BE	465	10	39	3.43	0
15. TENA_CHICK	TENASCIN PRECURSOR (TN) (HEXA	1808	10	36	3.43	0
**** 2 standard deviations above mean ****						
16. KR2_HUMAN	HKR2 PROTEIN (FRAGMENT).	106	9	23	2.86	0
17. WAP_RABIT	WHEY ACIDIC PROTEIN PRECURSOR	127	9	16	2.86	0
18. NU6M_DROYA	NADH-UBIQUINONE OXIDOREDUCTAS	174	9	15	2.86	0
19. NU6M_DROME	NADH-UBIQUINONE OXIDOREDUCTAS	174	9	16	2.86	0
20. YEIB_ECOLI	HYPOTHETICAL PROTEIN IN GALS	196	9	28	2.86	0
21. HBG3_MOUSE	INT-2 PROTO-ONCOGENE PROTEIN	245	9	34	2.86	0
22. NIFC_CLOPA	NIFC PROTEIN.	286	9	16	2.86	0
23. YCE9_YEAST	HYPOTHETICAL 35.6 KD PROTEIN	312	9	16	2.86	0
24. ASG2_ECOLI	L-ASPARAGINASE II PRECURSOR (	348	9	39	2.86	0
25. HC_HUMAN	ALPHA-1-MICROGLOBULIN / INTER	352	9	34	2.86	0
26. DBDR_RAT	D(1B) DOPAMINE RECEPTOR.	475	9	18	2.86	0
27. D5DR_HUMAN	D(5) DOPAMINE RECEPTOR.	477	9	19	2.86	0
28. LMP2_EBV	GENE TERMINAL PROTEIN (MEMBRA	497	9	37	2.86	0
29. MPP1_NEUCR	MITOCHONDRIAL PROCESSING PEPT	577	9	21	2.86	0
30. HS75_YEAST	HEAT SHOCK PROTEIN SSB1.	613	9	35	2.86	0
31. EF3_PNECA	ELONGATION FACTOR 3 (EF-3).	1042	9	40	2.86	0
32. NRG_DROME	NEUROGLIAN PRECURSOR.	1239	9	37	2.86	0
33. IP3R_DROME	INOSITOL 1,4,5-TRISPHOSPHATE-	2833	9	35	2.86	0
34. DEF1_RABIT	CORTICOSTATIN I PRECURSOR (CS	93	8	18	2.29	0
35. CYB_GEDSD	CYTOCHROME B (EC 1.10.2.2) (F	96	8	10	2.29	0
36. APC2_CAVPO	APOLIPOPROTEIN C-II PRECURSOR	100	8	15	2.29	0
37. VPX_HIV2D	VPX PROTEIN (X DRF PROTEIN).	111	8	19	2.29	0
38. VPX_SIVS4	VPX PROTEIN (X DRF PROTEIN).	112	8	19	2.29	0
39. COL_CANFA	COLIPASE PRECURSOR.	112	8	19	2.29	0
40. YSCB_VEREN	HYPOTHETICAL YSC OPERON PROTE	137	8	21	2.29	0

# 1. ELLIS-012-FIG2AB.PEP (1-256)

41BB\_MOUSE T CELL ANTIGEN 4-1BB PRECURSOR.

ID 41BB\_MOUSE STANDARD; PRT; 256 AA.  
AC P20334;  
DT 01-FEB-1991 (REL. 17, CREATED)  
DT 01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)  
DT 01-APR-1993 (REL. 25, LAST ANNOTATION UPDATE)  
DE T CELL ANTIGEN 4-1BB PRECURSOR.  
DS MUS MUSCULUS (MOUSE).  
DC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
DC EUTHERIA; RODENTIA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RM 89184547  
RA KWON B.S., WEISSMAN S.M.;  
RL PROC. NATL. ACAD. SCI. U.S.A. 86:1963-1967(1989).  
RN [2]  
RP CHARACTERIZATION, AND SEQUENCE OF 25-29.  
RA POLLOK K.E., KIM Y.-J., ZHOU Z., HURTADO J., KIN K.K., PICKARD R.T.,  
RA KWON B.S.;  
RL J. IMMUNOL. 150:771-781(1993).  
CC -!- FUNCTION: PUTATIVE RECEPTOR FOR A CYTOKINE. POSSIBLY ACTIVE  
CC DURING T CELL ACTIVATION.  
CC -!- SUBUNIT: PRINCIPALLY AN HOMODIMER, BUT ALSO FOUND AS A MONOMER.  
CC -!- INDUCTION: OPTIMAL BY PMA AND IONOMYCIN.  
CC -!- TISSUE SPECIFICITY: EXPRESSED ON THE SURFACE OF ACTIVATED T CELLS.  
CC -!- SIMILARITY: CONTAINS A LA-NGFR/TNFR-TYPE CYSTEINE-RICH REGION.  
DR FMRI: J04492; MMT41BB.

DR PIR; B32393; B32393.  
 DR PROSITE; PS00652; TNFR\_NGFR.  
 KW RECEPTOR; GLYCOPROTEIN; SIGNAL.  
 FT SIGNAL 1 24  
 FT CHAIN 25 256 T CELL ANTIGEN 4-1BB.  
 FT DOMAIN 17 159 NGFR/TNFR REPEATS.  
 FT REPEAT 17 45 NGFR/TNFR REPEAT 1.  
 FT REPEAT 46 85 NGFR/TNFR REPEAT 2.  
 FT REPEAT 86 117 NGFR/TNFR REPEAT 3.  
 FT REPEAT 118 159 NGFR/TNFR REPEAT 4.  
 FT CARBOHYD 128 128 POTENTIAL.  
 FT CARBOHYD 138 138 POTENTIAL.  
 SQ SEQUENCE 256 AA; 27598 MW; 347415 CN;

Initial Score = 256 Optimized Score = 256 Significance = 143.96  
 Residue Identity = 100% Matches = 256 Mismatches = 0  
 Gaps = 0 Conservative Substitutions = 0

```

X      10      20      30      40      50      60      70
MGNNCYNVVVIVLLLVGCEKVGAVQNSCDNCQPGTFCRKYNPVCKSCPPSTFSSIGGQPNCNICRVCAGYFR
|||||
MGNNCYNVVVIVLLLVGCEKVGAVQNSCDNCQPGTFCRKYNPVCKSCPPSTFSSIGGQPNCNICRVCAGYFR
X      10      20      30      40      50      60      70

      80      90     100     110     120     130     140
FKKFCSSSTHNAECEIEGFHCLGPQCTRCEKDCRPGQELTKQGCKTCSLGTFNQNGTGVCRPWTNCSLDGR
|||||
FKKFCSSSTHNAECEIEGFHCLGPQCTRCEKDCRPGQELTKQGCKTCSLGTFNQNGTGVCRPWTNCSLDGR
      80      90     100     110     120     130     140

     150     160     170     180     190     200     210
SVLKTGTTEKDVVCGPPVVSFSPSTTISVTPEGGPGGHSLOVLTFLALTSALLLALIFITLLFSVLKWIRK
|||||
SVLKTGTTEKDVVCGPPVVSFSPSTTISVTPEGGPGGHSLOVLTFLALTSALLLALIFITLLFSVLKWIRK
     150     160     170     180     190     200     210

    220     230     240     250     X
KFPHFIFKQPFKKTTGAAQEEEDACSCRCPOEEEGGGGYEL
|||||
KFPHFIFKQPFKKTTGAAQEEEDACSCRCPOEEEGGGGYEL
    220     230     240     250     X

```

## 2. ELLIS-012-FIG2AB.PEP (1-256)

TF\_MOUSE TISSUE FACTOR PRECURSOR (TF) (COAGULATION FACTOR I

ID TF\_MOUSE STANDARD; PRT; 294 AA.  
 AC P20352;  
 DT 01-FEB-1991 (REL. 17, CREATED)  
 DT 01-AUG-1991 (REL. 19, LAST SEQUENCE UPDATE)  
 DT 01-MAY-1992 (REL. 22, LAST ANNOTATION UPDATE)  
 DE TISSUE FACTOR PRECURSOR (TF) (COAGULATION FACTOR III).  
 GN CF-3.  
 OS MUS MUSCULUS (MOUSE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; RODENTIA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RM 91093171  
 RA RANGANATHAN G., BLATTI S.P., SUBRAMANIAM M., FASS D.N., MAIHLE N.J.,  
 RA GETZ M.J.;  
 RL J. BIOL. CHEM. 266:496-501(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BALB/C;  
 RM 89343974

RA HARTZELL S., RYDER K., LANAHAN A., LAU L.F., NATHANS D.;  
 RL MOL. CELL. BIOL. 9:2567-2573(1989).  
 CC -!- FUNCTION: INITIATES BLOOD COAGULATION BY FORMING A COMPLEX WITH  
 CC CIRCULATING FACTOR VII OR VIIA. THE [TF:VIIA] COMPLEX ACTIVATES  
 CC FACTORS IX OR X BY SPECIFIC LIMITED PROTOLYSIS. TF PLAYS A ROLE IN  
 CC NORMAL HEMOSTASIS BY INITIATING THE CELL-SURFACE ASSEMBLY AND  
 CC PROPAGATION OF THE COAGULATION PROTEASE CASCADE.  
 DR EMBL; M57896; MMTFA.  
 DR EMBL; M26071; MMTF.  
 DR PIR; A32318; A32318.  
 DR PIR; A39046; A39046.  
 DR PROSITE; PS00621; TISSUE\_FACTOR.  
 KW GLYCOPROTEIN; BLOOD COAGULATION; TRANSMEMBRANE; SIGNAL; LIPOPROTEIN.  
 FT SIGNAL 1 28  
 FT CHAIN 29 294 TISSUE FACTOR.  
 FT DOMAIN 29 251 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 252 274 POTENTIAL.  
 FT DOMAIN 275 294 CYTOPLASMIC (POTENTIAL).  
 FT SITE 245 247 WKS MOTIF.  
 FT CARBOHYD 37 37 POTENTIAL.  
 FT CARBOHYD 57 57 POTENTIAL.  
 FT CARBOHYD 169 169 POTENTIAL.  
 FT CARBOHYD 200 200 POTENTIAL.  
 FT DISULFID 75 83 BY SIMILARITY.  
 FT DISULFID 218 241 BY SIMILARITY.  
 FT LIPID 275 275 PALMITATE (BY SIMILARITY).  
 FT CONFLICT 26 26 I -> T (IN REF. 2).  
 SQ SEQUENCE 294 AA; 32935 MW; 468130 CN;

Initial Score = 11 Optimized Score = 42 Significance = 4.00  
 Residue Identity = 23% Matches = 58 Mismatches = 165  
 Gaps = 28 Conservative Substitutions = 0

```

                                X      10      20
                                MGNNCYNVVVIVLLVGCEK-V
                                | | |
VRPRLAALAPTFLGCLLLQVIAGAGIPEKAFNLTWISTDFKTIIEWQPKPTNYTYTVQISDRSRNWKNKCF
 10      20      30      40      50      X 60      70

      30      40      50      60      70      80      90
GAVQNSCDNCQPGTFCRKYNPVCK-SCPPSTFSSIGGQPNCNICRVCAGYFRFKKF-CSSTHNAECECIEGF
  ||      | | | | | | | | | | | | | | | | | |
STTDTECDLTDEIVKDVWYAEAKVLSVPRRNSVHGDQDLVIHGEEPPFTNAPKFLPYRDTNLGQPVIGQF
 80      90      100      110      120      130      140

      100      110      120      130      140      150
HCLGPQCTRCEKDCRPGQELTKQGCKTCSLGTFNQNG--TGVCRPHTNCSLDGRSVLKTGTTE--KDVVCG
  |      | | | | | | | | | | | | | | | | | |
EQDGRKLNVVVKD-----SLT-LVRKNGTFLTLRQVFGKDLGYIITYRKGSSTGKKTNITNTNEFSIDVEEG
150      160      170      180      190      200      210

160      170      180      190      200      210      220
PPVVSFSPSTTIS-VTPEGPGGHSLOVLT----LFLALT-----SALLLALIFITLLFSVLKWIRKKFPHI
  |      | | | | | | | | | | | | | | | | | |
VSYCFFVQAMIFSRKTQNSPG--SSTVCTEQWKSFLGETLIIVGAVVLLATIFIILLSISLCKRRK---NR
 220      230      240      250      260      270      280

      230      X 240      250
FKQPFKKTGAAQEEEDACSCRCPOEEEGGGGYEL
  | | |
AGQKGNTPSRLA
      290      X
  
```

ID NU5M\_ARTSX STANDARD; PRT; 59 AA.  
AC P19047;  
DT 01-NOV-1990 (REL. 16, CREATED)  
DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)  
DT 01-NOV-1990 (REL. 16, LAST ANNOTATION UPDATE)  
DE NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 5 (EC 1.6.5.3) (FRAGMENT).  
GN ND5.  
OS ARTEMIA SP. (BRINE SHRIMP).  
OG MITOCHONDRION.  
OC EUKARYOTA; METAZOA; ARTHROPODA; CRUSTACEA; BRANCHIOPODA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RM 88289417  
RA BATUECAS B., GARESSE R., CALLEJA M., VALVERDE J.R., MARCO R.;  
RL NUCLEIC ACIDS RES. 16:6515-6529(1988).  
CC -!- CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+) + UBIQUINOL.  
DR EMBL; X07663; M1AS07.  
DR PIR; S01877; S01877.  
KW OXIDOREDUCTASE; NAD; UBIQUINONE; MITOCHONDRION.  
FT NON\_TER 1 1  
SQ SEQUENCE 59 AA; 6585 MW; 22406 CN;

Initial Score = 10 Optimized Score = 14 Significance = 3.43  
Residue Identity = 25% Matches = 15 Mismatches = 42  
Gaps = 2 Conservative Substitutions = 0

110	120	130	140	150	160	170	180
ELTKQGCKTCSLGTDFNDQNGTGVCPRWTNCSLDGRSVLKTGTTEKDVVCGPPVVSFSPSTTISVTPEGGPGG							
						MGELLYHEGDCGWVEEAGPSLI	
					X	10	20
	190	200	210	X 220	230	240	250
H--SLQVLTFLALTSALLLALIFITLLFSVLKWKIRKKFPHIFKQPFKKTGAAQEEEDACSCRCPEEEEGGG							
HHNSLRGSSSLFSFLTSSPYKVLILSSLLFTLFMYMSA							
	30	40	50	X			

GGYEL

#### 4. ELLIS-012-FIG2AB.PEP (1-256)

M169\_MOUSE M1/69-J11D HEAT STABLE ANTIGEN PRECURSOR.

ID M169\_MOUSE STANDARD; PRT; 76 AA.  
AC P24807;  
DT 01-MAR-1992 (REL. 21, CREATED)  
DT 01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)  
DT 01-AUG-1992 (REL. 23, LAST ANNOTATION UPDATE)  
DE M1/69-J11D HEAT STABLE ANTIGEN PRECURSOR.  
GN HSA-A.  
OS MUS MUSCULUS (MOUSE).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; RODENTIA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RM 90361906  
RA KAY R., TAKEI F., HUMPHRIES R.K.;  
RL J. IMMUNOL. 145:1952-1959(1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CBA X C57BL/6; TISSUE=SPLEEN;  
RM 91209380  
RA WENCER R.H., AVANE M., ROSE P., KREHLER C., NIELSEN P.J.;

RL EUR. J. IMMUNOL. 21:1039-1046(1991).  
 CC -!- FUNCTION: MAY HAVE A SPECIFIC ROLE TO PLAY IN EARLY THYMOCYTE  
 CC DEVELOPMENT.  
 CC -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.  
 CC -!- TISSUE SPECIFICITY: IN LYMPHOID, MYELOID, AND ERYTHROID CELLS.  
 CC -!- SIMILARITY: TO HUMAN SIGNAL TRANSDUCER CD24.  
 DR EMBL; M58661; M169J1.  
 DR EMBL; X56469; MMHSAAG.  
 DR PIR; S15784; S15784.  
 DR PIR; A43537; A43537.  
 KW ANTIGEN; SIGNAL; GPI-ANCHOR; GLYCOPROTEIN; MULTIGENE FAMILY; MEMBRANE.  
 FT SIGNAL 1 26 POTENTIAL.  
 FT CHAIN 27 56 M1/69-J11D ANTIGEN.  
 FT PROPEP 57 76 REMOVED IN MATURE FORM (POTENTIAL).  
 FT LIPID 56 56 GPI-ANCHOR (POTENTIAL).  
 FT CARBOHYD 27 27 POTENTIAL.  
 FT CARBOHYD 39 39 POTENTIAL.  
 FT CARBOHYD 48 48 POTENTIAL.  
 SQ SEQUENCE 76 AA; 7797 MW; 30445 CN;

Initial Score = 10 Optimized Score = 19 Significance = 3.43  
 Residue Identity = 28% Matches = 22 Mismatches = 52  
 Gaps = 4 Conservative Substitutions = 0

```

      80      90      100      110      120      130      140
SSTHNAECECIEGFHCLGPQCTRCEKDCRPGQELTKQCKTCSLGTENDQNGTGVCRPWTNCSLDGRSVLKT
                                     |         |         |
                                     MGRAMVARLGLGLLLALLLPT
                                     X          10         20

150      160      170      180      190      200 X  210
--GTTEKDVVCGPPVVSFSPSTTISVTPEGGPGHSLQVLTFLALTSALLLALIFITLLFSVLKWIRKKFP
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
GIYCNQTSVAPFPGNQISASPNPSNATTRG-GGSSLQSTAGLLAL-SLSLLHLYC
      30      40      50      60      70      X

220      230      240      250
HIFKQPFKKTGAAQEEEDACSCRCPEEEEGGGG
  
```

##### 5. ELLIS-012-FIG2AB.PEP (1-256)

ZEAB\_MAIZE ZEIN-ALPHA PRECURSOR (19 KD) (CLONE PZ19.1) (FRAGM

ID ZEAB\_MAIZE STANDARD; PRT; 186 AA.  
 AC PD4705;  
 DT 13-AUG-1987 (REL. 05, CREATED)  
 DT 13-AUG-1987 (REL. 05, LAST SEQUENCE UPDATE)  
 DT 01-AUG-1988 (REL. 08, LAST ANNOTATION UPDATE)  
 DE ZEIN-ALPHA PRECURSOR (19 KD) (CLONE PZ19.1) (FRAGMENT).  
 OS ZEA MAYS (MAIZE).  
 OC EUKARYOTA; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; MONOCOTYLEDONEAE;  
 OC CYPERALES; GRAMINEAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RM 83103094  
 RA PEDERSEN K., DEVEREUX J., WILSON D.R., SHELDON E., LARKINS B.A.;  
 RL CELL 29:1015-1026(1982).  
 CC -!- FUNCTION: ZEINS ARE MAJOR SEED STORAGE PROTEINS.  
 CC -!- THE ALPHA ZEINS OF 19 KD AND 22 KD ACCOUNT FOR 70% OF THE TOTAL  
 CC ZEIN FRACTION. THEY ARE ENCODED BY A LARGE MULTIGENE FAMILY.  
 DR EMBL; V01471; ZMZED2.  
 KW SEED STORAGE PROTEIN; TANDEM REPEAT; MULTIGENE FAMILY; SIGNAL.  
 FT SIGNAL 1 21  
 FT CHAIN 22 >186 ZEIN-ALPHA.  
 FT NON\_TER 186 186  
 SQ SEQUENCE 186 AA; 20419 MW; 154904 CN;

70 80 90 100 110 X 120 130  
ICRV CAGYFRFKKCSSTHNAECEIEGFHCLGPQCTRCEKDCRPGQELTKQGCKTCSLGT FNDQNGTGVC R  
| | | | |  
KTECEIMLC-LASAAATATIE

```

140      150      160      170      180      190      200
PWTNCSLDGRSVLKTGTTEKDV--VCGPPVVSFSPSTTISVTPEGGPGGHSLOVLTFLALTSALLLALIFI
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
P--QCSQAPITSLPPYLSPAVSSVCENP--ILQPYRIQQAIAAG-----ILPLSPLFLQGPSALLQQLPLV
      30      40      50      60      70      80

```

```

210      220      230      240      250      X
TLLFSVLKWIRKKFPHIFKQPFKKTGAQEEDACSCRCPOEEEGGGGGYEL
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
HLL---AQNIR----AQQQLQGLVLGNLAAYSQGHQFLPFNGLAALNSAAYLQQLPFSSQLAAAYPQQFLPFN
      90      100      110      120      130      140

```

```

GLAALNSAAYLQQLPFSQLADVSPAAF
150      160      170

```

## 7. ELLIS-012-FIG2AB.PEP (1-256)

ZEAS\_MAIZE ZEIN-ALPHA PRECURSOR (19 KD) (CLONE GZ19AB11).

```

ID  ZEAS_MAIZE      STANDARD;      PRT;   234 AA.
AC  P08416;
DT  01-AUG-1988 (REL. 08, CREATED)
DT  01-AUG-1988 (REL. 08, LAST SEQUENCE UPDATE)
DT  01-AUG-1988 (REL. 08, LAST ANNOTATION UPDATE)
DE  ZEIN-ALPHA PRECURSOR (19 KD) (CLONE GZ19AB11).
DS  ZEA MAYS (MAIZE).
OC  EUKARYOTA; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; MONOCOTYLEDONEAE;
DC  CYPERALES; GRAMINEAE.
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=W64A;
RM  87257300
RA  KRIZ A.L., BOSTON R.S., LARKINS B.A.;
RL  MOL. GEN. GENET. 207:90-98(1987).
CC  -!- FUNCTION: ZEINS ARE MAJOR SEED STORAGE PROTEINS.
CC  -!- THE ALPHA ZEINS OF 19 KD AND 22 KD ACCOUNT FOR 70% OF THE TOTAL
CC      ZEIN FRACTION. THEY ARE ENCODED BY A LARGE MULTIGENE FAMILY.
DR  EMBL; X05911; ZMZEI19.
DR  PIR; S03417; S03417.
KW  SEED STORAGE PROTEIN; TANDEM REPEAT; SIGNAL.
FT  SIGNAL          1      21
FT  CHAIN          22      234      ZEIN-ALPHA.
SQ  SEQUENCE      234 AA;  25439 MW;  271676 CN;

```

```

Initial Score      =      10  Optimized Score =      23  Significance =  3.43
Residue Identity   =     21%  Matches          =      32  Mismatches   =     98
Gaps               =      19  Conservative Substitutions =      0

```

```

60      70      80      90      100      110      120      130
NCNICRVCAGYFRFKKFCSSHTNAECECIEGFHCLGPQCTRCEKDCRPGQELTKQGCKTCSLGTFFNDQNGTG
                                     |  |  |
                                     MAAKIFCLLMLLG---LSASAA
                                     X      10

```

```

140      150      160      170      180      190      200
VCRPWTNCSLDGRSVLKTGTTEKDV--VCGPPVVSFSPSTTISVTPEGGPGGHSLOVLTFLALTSALLLAL
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
TATIFTQCSQAPIASLLPPYLSAVSSVCENP--ILQPYRIQQAIAAG-----ILPLSPLFLQSSALLQQL
20      30      40      50      60      70      80

```

```

210      220      230      240      250      X
IFITLLFSVLKWIRKKFPHIFKQPFKKTGAQEEDACSCRCPOEEEGGGGGYEL
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
PIVHI---AQNIR----AQQQLQGLVLGNLAAYSQGHQFLPFNGLAALNSAAYLQQLPFSQLAAAYPQQFLPFN

```

LPFNQLAALNSPAYLQQQQLLPFSQLAGVSPAT  
150 160 170 180

## 8. ELLIS-012-FIG2AB.PEP (1-256)

ZEA4\_MAIZE ZEIN-ALPHA PRECURSOR (19 KD) (CLONE 19B1).

ID ZEA4\_MAIZE STANDARD; PRT; 234 AA.  
AC P06675;  
DT 01-JAN-1988 (REL. 06, CREATED)  
DT 01-JAN-1988 (REL. 06, LAST SEQUENCE UPDATE)  
DT 01-NOV-1988 (REL. 09, LAST ANNOTATION UPDATE)  
DE ZEIN-ALPHA PRECURSOR (19 KD) (CLONE 19B1).  
OS ZEA MAYS (MAIZE).  
OC EUKARYOTA; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; MONOCOTYLEDONEAE;  
OC CYPERALES; GRAMINEAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RM 86059563  
RA MARKS M.D., LINDELL J.S., LARKINS B.A.;  
RL J. BIOL. CHEM. 260:16451-16459(1985).  
CC -!- FUNCTION: ZEINS ARE MAJOR SEED STORAGE PROTEINS.  
CC -!- THE ALPHA ZEINS OF 19 KD AND 22 KD ACCOUNT FOR 70% OF THE TOTAL  
CC ZEIN FRACTION. THEY ARE ENCODED BY A LARGE MULTIGENE FAMILY.  
CC -!- STRUCTURALLY, 22K AND 19K ZEINS ARE COMPOSED OF NINE ADJACENT,  
CC TOPOLOGICALLY ANTIPARALLEL HELICES CLUSTERED WITHIN A DISTORTED  
CC CYLINDER.  
DR EMBL; M12143; ZMZE19B1.  
DR PIR; E24557; ZIZMB1.  
KW SEED STORAGE PROTEIN; TANDEM REPEAT; MULTIGENE FAMILY; SIGNAL.  
FT SIGNAL 1 21  
FT CHAIN 22 234 ZEIN-ALPHA.  
SQ SEQUENCE 234 AA; 25435 MW; 271626 CN;

Initial Score = 10 Optimized Score = 23 Significance = 3.43  
Residue Identity = 22% Matches = 33 Mismatches = 97  
Gaps = 19 Conservative Substitutions = 0

60 70 80 90 100 110 120 130  
NCNICRVCAGYFRFKKFCSSHTNAECECIEGFHCLGPQCTRCEKDCRPGQELTKQGCKTCSLGTENDQNGTG  
| | | |  
MAAKIFCLMLLG-LSASAATA  
X 10 20

140 150 160 170 180 190 200  
VCRPWNTCSLDGRSVLKTGTTEKDV--VCGPPVVSFSPSTTISVTPEGPGGHSLOVLTFLALTSALLLAL  
| | | | | | | | | |  
TIFP--QCSQAPIASLLPPYLSSAVSSVCENP--ILQPYRIQQAIAAG-----ILPLSPLFLQSSALLQQL  
30 40 50 60 70 80

210 220 230 240 250 X  
IFITLLFSVLKWKIRKFKPHIFKQPFKKTGAAQEEEDACSCRCPEEEGGGGGYEL  
| | | | |  
PLVHLL---AQNIR----AQQLQQLVLANLAAYSQQQQFLPFNLGSLNSASYLQQQQLPFSQLPAAYPQQF  
90 100 110 120 130 X 140

LPFNQLAALNSPAYLQQQQLLPFSQLAGVSPAT  
150 160 170 180

## 9. ELLIS-012-FIG2AB.PEP (1-256)

ZEA1\_MAIZE ZEIN-ALPHA PRECURSOR (19 KD) (CLONE A30).

ID ZEA1\_MAIZE STANDARD; PRT; 234 AA



AC P02859;  
 DT 21-JUL-1986 (REL. 01, CREATED)  
 DT 13-AUG-1987 (REL. 05, LAST SEQUENCE UPDATE)  
 DT 01-AUG-1988 (REL. 08, LAST ANNOTATION UPDATE)  
 DE ZEIN-ALPHA PRECURSOR (19 KD) (CLONE A30).  
 OS ZEA MAYS (MAIZE).  
 OC EUKARYOTA; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; MONOCOTYLEDONEAE;  
 OC CYPERALES; GRAMINEAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RM 82081837  
 RA GERAGHTY D., PEIFER M.A., RUBENSTEIN I., MESSING J.;  
 RL NUCLEIC ACIDS RES. 9:5163-5174(1981).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RM 84207882  
 RA HU N.T., PEIFER M.A., HEIDECKER G., MESSING J., RUBENSTEIN I.;  
 RL EMBO J. 1:1337-1342(1982).  
 CC -!- FUNCTION: ZEINS ARE MAJOR SEED STORAGE PROTEINS.  
 CC -!- THE ALPHA ZEINS OF 19 KD AND 22 KD ACCOUNT FOR 70% OF THE TOTAL  
 CC ZEIN FRACTION. THEY ARE ENCODED BY A LARGE MULTIGENE FAMILY.  
 CC -!- STRUCTURALLY, 22K AND 19K ZEINS ARE COMPOSED OF NINE ADJACENT,  
 CC TOPOLOGICALLY ANTIPARALLEL HELICES CLUSTERED WITHIN A DISTORTED  
 CC CYLINDER.  
 DR EMBL; V01481; ZMZEIN.  
 DR PIR; C22762; Z1ZM3.  
 DR PIR; S21970; S21970.  
 KW SEED STORAGE PROTEIN; TANDEM REPEAT; MULTIGENE FAMILY; SIGNAL.  
 FT SIGNAL 1 21  
 FT CHAIN 22 234 ZEIN-ALPHA.  
 SQ SEQUENCE 234 AA; 25403 MW; 260041 CN;

Initial Score = 10 Optimized Score = 23 Significance = 3.43  
 Residue Identity = 22% Matches = 33 Mismatches = 97  
 Gaps = 19 Conservative Substitutions = 0

```

60      70      80      90     100     110     120     130
NCNICRVCAGYFRFKKFCSSTHNAECECIEGFHCLGPQCTRCEKDCRPGQELTKQGCKTCSLGTfNDQNGTG
                                     | | | |
                                     MAAKIFCLLMLLG-LSASAATA
                                     X      10      20

140     150     160     170     180     190     200
VCRPWTNCSLDGRSVLKTGTTEKDV--VCGPPVVVSFSPSTTISVTPEGGPGGHSLOVLTFLALTSALLLAL
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
TIFF--QCSQAPIASLLPPVLSPAVSSVCENP--ILQPYRIQQAIAAG-----ILPLSPLFLQSSALLQQL
      30      40      50      60      70      80

210     220     230     240     250     X
IFITLLFSVLKWKIRKFKPHIFKQPFKKTGAAQEEDACSCRCPQEEEGGGGYEL
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
PLVHLL---AGNIR----AQQLQQLVLANLAAYSQQQQFLPFNQLAALNSASYLQQQQLPFSQLPAAYPQQF
      90     100     110     120     130 X     140

LPFNQLAALNSPAYLQQQQLLPFSQLAGVSPAT
150     160     170     180
  
```

10. ELLIS-012-FIG2AB.PEP (1-256)  
 ZEAC\_MAIZE ZEIN-ALPHA PRECURSOR (19 KD) (PMS1).

ID ZEAC\_MAIZE STANDARD; PRT; 235 AA.  
 AC P24449;  
 DT 01-MAR-1992 (REL. 21, CREATED)  
 DT 01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)  
 DT 01-MAR-1992 (REL. 21, LAST ANNOTATION UPDATE)

DE ZEIN-ALPHA PRECURSOR (19 KD) (PMS1).  
 GN ZMPMS1.  
 OS ZEA MAYS (MAIZE).  
 OC EUKARYOTA; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; MONOCOTYLEDONEAE;  
 DC CYPERALES; GRAMINEAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. A619;  
 RM 90060774  
 RA QUAYLE T.J.A., BROWN J.W.S., FEIX G.;  
 RL GENE 80:249-257(1989).  
 CC -!- FUNCTION: ZEINS ARE MAJOR SEED STORAGE PROTEINS.  
 CC -!- THE ALPHA ZEINS OF 19 KD AND 22 KD ACCOUNT FOR 70% OF THE TOTAL  
 CC ZEIN FRACTION. THEY ARE ENCODED BY A LARGE MULTIGENE FAMILY.  
 CC -!- STRUCTURALLY, 22K AND 19K ZEINS ARE COMPOSED OF NINE ADJACENT,  
 CC TOPOLOGICALLY ANTIPARALLEL HELICES CLUSTERED WITHIN A DISTORTED  
 CC CYLINDER.  
 DR EMBL; X53582; ZMPMS1G.  
 DR PIR; S15655; S15655.  
 KW SEED STORAGE PROTEIN; TANDEM REPEAT; MULTIGENE FAMILY; SIGNAL.  
 FT SIGNAL 1 21 BY SIMILARITY.  
 FT CHAIN 22 235 ZEIN-ALPHA.  
 SQ SEQUENCE 235 AA; 25505 MW; 262683 CN;

Initial Score = 10 Optimized Score = 23 Significance = 3.43  
 Residue Identity = 22% Matches = 33 Mismatches = 97  
 Gaps = 19 Conservative Substitutions = 0

```

60      70      80      90     100     110     120     130
NCNICRVCAGYFRFKKFCSSSTHNAECECIEGFHCLGPQCTRCEKDCRPGQELTKQGCKTCSLGTENDQNGTG
                                     | |  | |
                                     MAAKIFCLLMLLG-LSASAATA
                                     X      10      20

140      150      160      170      180      190      200
VCRPWTNCSLDGRSVLKTGTTEKDV--VCGPPVVSFSPSTTISVTPEGGPGGHSLOQLTLFLALTSALLLAL
|  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
TIFP--QCSQAPIASLLPPYLSPAVSSVCENP--ILQPYRIQQAIAAG-----ILPLSPLFLQSSALLQQL
      30      40      50      60      70      80

210      220      230      240      250      X
IFITLLFSVLKWKIRKKFPHIFKQPFKKTGAAQEEADACSCRCPEEEGGGGGYEL
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
PLVHLL---AQNIR----AQQLQQLVLANVAAYSQGGQFLPFNQLAALNSAAYLQGGQLLPFSQLTAAYPQQ
      90      100      110      120      130 X      140

FLPFNQLAALNSAAYLQGGQLLPFSQLAVVSPA
150      160      170      180
  
```

11. ELLIS-012-FIG2AB.PEP (1-256)  
 ZEA2\_MAIZE ZEIN-ALPHA PRECURSOR (19 KD) (CLONE ZG99).

ID ZEA2\_MAIZE STANDARD; PRT; 235 AA.  
 AC P04704;  
 DT 13-AUG-1987 (REL. 05, CREATED)  
 DT 13-AUG-1987 (REL. 05, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1988 (REL. 09, LAST ANNOTATION UPDATE)  
 DE ZEIN-ALPHA PRECURSOR (19 KD) (CLONE ZG99).  
 OS ZEA MAYS (MAIZE).  
 OC EUKARYOTA; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; MONOCOTYLEDONEAE;  
 DC CYPERALES; GRAMINEAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RM 82265740  
 RA MARKS M D , LARKINS R A

RL J. BIOL. CHEM. 257:9976-9983(1982).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RM 83103094  
 RA PEDERSEN K., DEVEREUX J., WILSON D.R., SHELDON E., LARKINS B.A.;  
 RL CELL 29:1015-1026(1982).  
 CC -!- FUNCTION: ZEINS ARE MAJOR SEED STORAGE PROTEINS.  
 CC -!- THE ALPHA ZEINS OF 19 KD AND 22 KD ACCOUNT FOR 70% OF THE TOTAL  
 CC ZEIN FRACTION. THEY ARE ENCODED BY A LARGE MULTIGENE FAMILY.  
 CC -!- STRUCTURALLY, 22K AND 19K ZEINS ARE COMPOSED OF NINE ADJACENT,  
 CC TOPOLOGICALLY ANTIPARALLEL HELICES CLUSTERED WITHIN A DISTORTED  
 CC CYLINDER.  
 DR EMBL; V01470; ZMZE01.  
 DR EMBL; V01479; ZMZE10.  
 DR PIR; A29288; Z1ZM99.  
 KW SEED STORAGE PROTEIN; TANDEM REPEAT; MULTIGENE FAMILY; SIGNAL.  
 FT SIGNAL 1 21  
 FT CHAIN 22 235 ZEIN-ALPHA.  
 SQ SEQUENCE 235 AA; 25575 MW; 261593 CN;

Initial Score = 10 Optimized Score = 22 Significance = 3.43  
 Residue Identity = 21% Matches = 32 Mismatches = 98  
 Gaps = 19 Conservative Substitutions = 0

```

60      70      80      90     100     110     120     130
NCNICRVCAGYFRFKKFCSSSTHNAECECIEGFHCLGPQCTRCEKDCRPGQELTKQGCKTCSLGTENDQNGTG
                                     | |  ||
                                     MAAKIFCLIMLLG-LSASAATA
                                     X      10      20

140      150      160      170      180      190      200
VCRPWTNCSLDGRSVL--KTGTTEKDVVCGPPVVSFSPSTTISVTPEGGPGGHSLOVLTFLALTSALLLAL
|  ||  |  ||  |  ||  |  ||  |  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
SIFP--QCSQAPIASLLPPYLSAMSSVCENP--ILLPYRIQQAIAAG-----ILPLSPLFLQSSALLQQL
      30      40      50      60      70      80

210      220      230      240      250      X
IFITLLFSVLKWKIRKKFPHIFKQPFKKTGAAQEEEDACSCRCPEEEEGGGGGYEL
||  ||  |  ||  |  ||  |  ||  |
PLVHLL---AQNIR----AQQLQQLVLANLAAYSQQQQFLPFNQLAALNSAAYLQQQQQLLPFSQLAAAYPRQ
      90      100      110      120      130 X      140

FLPFNQLAALNSHAYVQQQQQLLPFSQLAAVSPA
150      160      170      180
  
```

## 12. ELLIS-012-FIG2AB.PEP (1-256)

ZEAL\_MAIZE ZEIN-ALPHA PRECURSOR (CLONE Z4).

ID ZEAL\_MAIZE STANDARD; PRT; 253 AA.  
 AC P04701;  
 DT 13-AUG-1987 (REL. 05, CREATED)  
 DT 13-AUG-1987 (REL. 05, LAST SEQUENCE UPDATE)  
 DT 01-AUG-1988 (REL. 08, LAST ANNOTATION UPDATE)  
 DE ZEIN-ALPHA PRECURSOR (CLONE Z4).  
 OS ZEA MAYS (MAIZE).  
 OC EUKARYOTA; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; MONOCOTYLEDONEAE;  
 OC CYPERALES; GRAMINEAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RM 84207882  
 RA HU N.T., PEIFER M.A., HEIDECKER G., MESSING J., RUBENSTEIN I.;  
 RL EMBO J. 1:1337-1342(1982).  
 CC -!- FUNCTION: ZEINS ARE MAJOR SEED STORAGE PROTEINS.  
 CC -!- THE ALPHA ZEINS OF 19 KD AND 22 KD ACCOUNT FOR 70% OF THE TOTAL  
 CC ZEIN FRACTION. THEY ARE ENCODED BY A LARGE MULTIGENE FAMILY.

DR EMBL; V01472; ZMZE03.  
 KW SEED STORAGE PROTEIN; TANDEM REPEAT; MULTIGENE FAMILY; SIGNAL.  
 FT SIGNAL 1 21  
 FT CHAIN 22 253 ZEIN-ALPHA.  
 SQ SEQUENCE 253 AA; 27700 MW; 300631 CN;

Initial Score = 10 Optimized Score = 21 Significance = 3.43  
 Residue Identity = 19% Matches = 29 Mismatches = 105  
 Gaps = 15 Conservative Substitutions = 0

```

60      70      80      90     100     110     120     130
NCNICRVCAGYFRFKKFCSSTHNAECECIEGFHCLGPQCTRCEKDCRPGQELTKQGCKTCSLGTENDQNGTG
                                     | |  ||
                                     MAAKIFCLIMLLG-LSASAATA
                                     X      10      20

140      150      160      170      180      190      200
VCRPHTNCSLDGRSVL--KTGTTEKDVVCGPPVVSFSPSTTISVTPEGGPGGHSLOVLTFLALTSALLLAL
|  ||      |      || |      |      |      |      |||  |||| |
SIFP--QCSQAPIASLLPPYLSPAMSSVCENP--ILLPYRIQQAIAAG-----ILPLSPLFLQSSALLQQL
      30      40      50      60      70      80

210      220      230      240      250      X
IFITLLFSVLKWKIRKFKPHIFKQPFKKTGAAQEEADACSCRCPEEEGGGGGYEL
||      ||      |
PLVHLL---AQNIRAQQLQQLVLNLAAYSQQLPLVHLLAQNIRAQQLQQLVLNLAAYSQQLQQLFLPFNQ
90      100      110      120      130      X 140      150

LAALNSAAYLQQLLPFSQLAAAYPRQFLPFN
160      170      180

```

13. ELLIS-012-FIG2AB.PEP (1-256)  
 M034\_MOUSE MOV34 PROTEIN.

ID M034\_MOUSE STANDARD; PRT; 321 AA.  
 AC P26516;  
 DT 01-AUG-1992 (REL. 23, CREATED)  
 DT 01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)  
 DT 01-APR-1993 (REL. 25, LAST ANNOTATION UPDATE)  
 DE MOV34 PROTEIN.  
 GN MOV-34.  
 OS MUS MUSCULUS (MOUSE).  
 DC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; RODENTIA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RM 91005870  
 RA GRIDLEY T., GRAY D.A., ORR-WEAVER T., SORIANO P., BARTON D.E.,  
 RA FRANCKE U., JAENISCH R.;  
 RL DEVELOPMENT 109:235-242(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RM 92128931  
 RA GRIDLEY T., JAENISCH R., GENDRON-MAGUIRE M.;  
 RL GENOMICS 11:501-507(1991).  
 CC -!- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN EARLY DEVELOPMENT.  
 CC -!- DISEASE: DISRUPTION OF THE MOV-34 LOCUS IS A RECESSIVE EMBRYONIC  
 CC LETHAL MUTATION.  
 CC -!- SIMILARITY: 62% IDENTITY TO DROSOPHILA MOV34 PROTEIN.  
 DR EMBL; M64641; MMOV34.  
 DR EMBL; M64634; MMOV341.  
 DR EMBL; M64635; MMOV342.  
 DR EMBL; M64636; MMOV343.  
 DR EMBL; M64637; MMOV344.  
 DR EMBL; M64638; MMOV345.

DR EMBL; M64640; MMDV347.  
 DR PIR; A40556; BWMSV4.  
 FT DOMAIN 283 321 HYDROPHILIC.  
 SQ SEQUENCE 321 AA; 36540 MW; 520650 CN;

Initial Score = 10 Optimized Score = 38 Significance = 3.43  
 Residue Identity = 17% Matches = 46 Mismatches = 202  
 Gaps = 12 Conservative Substitutions = 0

```

      X      10      20      30      40
      MGNNCYNVVVIVLLLVGCEKVGAVQNSCDNCQPGTFCRKYNPVC
      ||      || ||      || ||      |      |      |
MPELAVQKVVVHPLVLLSVVDHFNRIKVGNN--QKRVVGVLLGSWQKKVLDVSN--FAVPFDEDDKDDSVW
      10      20      30      40      50      60

      50      60      70      80      90      100      110
      KSCPSTFSSIGGQPNCN-ICRVCAGYFRFKK-FCSSTHNAECECIEGFHCLGPQCTRCEKDCRPGQELTKQ
      |      |      |      |      |      |      |
      FLDHDYLENMYGNFKKVNARERIVGWYHTGPKLHKNDIAINELMKRYCPNSVLVIIDVKPKDLGLPTEAYIS
      70      80      90      100      110      120      130      140

      120      130      140      150      160      170      180
      GCKTCSLGTFDNGTGVCPRWTNCSLDGRSVLKTGTTEKDVVCGPPVVSFSPSTTISVTPEGGPGGHSLS-Q
      ||      |      |      |      |      |      |
      VEEVHDDGTPTSKTFEHVTSEIGAEAEVEVGVEHLLRDIKD----TTVGTLSQRITNQVHGLKGLNSKLLDI
      150      160      170      180      190      200

      190      200      210      220      230      240      250      X
      VLTFLALTSALLLALIFITLLFSVLKWKIRKFPKIF-KQPFKKTGAAGQEDACSCRCPEEEEGGGGGYEL
      |      |      |      |      |      |      |
      RSYLEKVASGKLPINHIIYQLQDVFNLLPDASLQEFVKAFYKLTNDQMVVVYLASLIRSVVALHNLINNKI
      210      220      230      240      250      260      270      280

      ANRDAEKKEGQKEESKKERKDDKEKEKSDAAKKEEKKKK
      290      300      310      320

```

#### 14. ELLIS-012-FIG2AB.PEP (1-256)

ATPB\_SULAC MEMBRANE-ASSOCIATED ATPASE BETA CHAIN (EC 3.6.1.34)

ID ATPB\_SULAC STANDARD; PRT; 465 AA.  
 AC P13052;  
 DT 01-JAN-1990 (REL. 13, CREATED)  
 DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)  
 DT 01-MAY-1992 (REL. 22, LAST ANNOTATION UPDATE)  
 DE MEMBRANE-ASSOCIATED ATPASE BETA CHAIN (EC 3.6.1.34) (SUL-ATPASE BETA).  
 GN ATPB.  
 OS SULFOLOBUS ACIDOCALDARIUS.  
 OC PROKARYOTA; MENDOSICUTES; ARCHAEBACTERIA; SULFOLOBALES.  
 RN [1]  
 RF SEQUENCE FROM N.A.  
 RM 89034240  
 RA DENDA K., KONISHI J., OSHIMA T., DATE T., YOSHIDA M.;  
 RL J. BIOL. CHEM. 263:17251-17254(1988).  
 CC -!- THIS IS A REGULATORY SUBUNIT.  
 CC -!- SUBUNIT: SUL-ATPASE IS COMPOSED OF SIX (OR FIVE ?) SUBUNITS:  
 CC ALPHA, BETA, DELTA, GAMMA, C (PROTEOLIPID), AND POSSIBLY EPSILON.  
 CC -!- SIMILARITY: STRONG TO OTHER ARCHAEBACTERIA BETA SUBUNITS, ALSO  
 CC RELATED TO THE ALPHA SUBUNITS OF FO-F1 ATPASES.  
 DR EMBL; M22402; SAATPB.  
 DR PIR; A32118; A32118.  
 DR PROSITE; PS00152; ATPASE\_ALPHA\_BETA.  
 KW HYDROLASE; HYDROGEN ION TRANSPORT.  
 SQ SEQUENCE 465 AA; 51247 MW; 1080510 CN;

Initial Score = 10 Optimized Score = 38 Significance = 3.43

Residue Identity = 17% Matches = 49 Mismatches = 204  
 Gaps = 21 Conservative Substitutions = 0

```

                                X      10      20      30
                                MGNNCYNVVVIVLLLVGCEKV--GAVQNSCDNC
                                |      ||      |      |      |
MSLLNVREYSNISMIGPLIAVQGVSDAAYNELVEIEMPDGSKRRGLVVDSDMGVTFVQVFEGTTGISPTGS
    10      20      30      40      50      60      70

    40      50      60      70      80      90      100
QPGTFCRKYNPVCKSCPPSTFSSIGGQPNPCNICRVCAGYFRFKKFCSSTHNAECECIEGFHCLGPQCTRCEK
    |      |      |      |      |      |      |      |      |
KVRFLGRGLEVKISEEMLGRIFNPLGEPLDNGPPVIGGEKR--NINGDPINPATREYPEEFIQTGISAIDGLN
    80      90      100      110      120      130      140

    110      120      130      140      150      160      170
DCRPGQELTKQCKTCSLGTFNQ--NGTGVCRPWTNCSLDGRSVLKTGTTEKDVVCGPPVVSFSPSTTISV
    |      |      |      |      |      |      |
SLLRGSKITDLSGSGLPANTLAAQIAKQATVRGEESNFAVVFAAIGVRYDEALFFRKFFETGAINRVAMFV
    150      160      170      180      190      200      210

    180      190      200      210      220      230
TPEGPGGHSLOVLTFLFALTSALLA-----LIFIT---LLFSVLKWIRKKFPHIFKQP-FKKTIG
    |      |      ||      ||      |||      ||      ||      |      |      |      |
TLANDP--PSLKILTPKTALTAEYLAFAEKDMHVLAILIDMTNYCEALRELSASREEVPGRGGYPGMYTDL
    220      230      240      250      260      270      280

```

```

    240      250      X
AAQEEDACSCRCPEEEEGGGGYEL
|  |  |
ATIERAGKVGKKGSITQMPILTMPNDDMTHPIPDLTGYITEGQIVLDRSLFNKGIYPPINVLMSLSRLMK
    290      300      310      320      330      340      350

```

DGI  
 360

# 15. ELLIS-012-FIG2AB.PEP (1-256)

TENA\_CHICK TENASCIN PRECURSOR (TN) (HEXABRACHION) (CYTOTACTIN)

```

ID  TENA_CHICK      STANDARD;      PRT;  1808 AA.
AC  P10039; P13132;
DT  01-MAR-1989 (REL. 10, CREATED)
DT  01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)
DT  01-AUG-1992 (REL. 23, LAST ANNOTATION UPDATE)
DE  TENASCIN PRECURSOR (TN) (HEXABRACHION) (CYTOTACTIN) (NEURONECTIN)
DE  (GMEM) (JI) (MIOTENDINOUS ANTIGEN) (GLIOMA-ASSOCIATED-EXTRACELLULAR
DE  MATRIX ANTIGEN) (CP 150-225).
OS  GALLUS GALLUS (CHICKEN).
OC  EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; AVES; NEOGNATHAE;
OC  GALLIFORMES.
RN  [1]
RP  SEQUENCE FROM N.A.
RC  TISSUE=EMBRYO;
RM  90030407
RA  SPRING J., BECK K., CHIQUET-EHRISMANN R.;
RL  CELL 59:325-334(1989).
RN  [2]
RP  SEQUENCE OF 27-722 FROM N.A., AND SEQUENCE OF 79-96.
RC  TISSUE=FIBROBLAST;
RM  89030589
RA  PEARSON C.A., PEARSON D., SHIBAHARA S., HOFSTEENGE J.,
RA  CHIQUET-EHRISMANN R.;
RL  EMBO J. 7:2977-2982(1988).
RN  [3]
RP  SEQUENCE OF 444-1018 AND 1412-1441 FROM N.A.

```

RM 88176910  
 RA JONES F.S., BURGOON M.P., HOFFMAN S., CROSSIN K.L., CUNNINGHAM B.A.,  
 RA EDELMAN G.M.;  
 RL PROC. NATL. ACAD. SCI. U.S.A. 85:2186-2190(1988).  
 CC -!- FUNCTION: SAM (SUBSTRATE-ADHESION MOLECULE) THAT APPEARS TO  
 CC INHIBIT CELL MIGRATION. MAY PLAY A ROLE IN SUPPORTING THE GROWTH  
 CC OF EPITHELIAL TUMORS.  
 CC -!- SUBUNIT: HEXAMERIC. AN HOMOTRIMER MAY BE FORMED IN THE TRIPLE  
 CC COILED-COIL REGION AND MAY BE STABILIZED BY DISULFIDE RINGS AT  
 CC BOTH ENDS. TWO OF SUCH HALF-HEXABRACHIONS MAY BE DISULFIDE LINKED  
 CC WITHIN THE CENTRAL GLOBULE.  
 CC -!- INDUCTION: BY TGF-BETA.  
 CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX.  
 CC -!- ALTERNATIVE SPLICING: THREE VARIANTS OF 230 KD, 200 KD, AND 190 KD  
 CC ARE PRODUCED FROM A SINGLE GENE IN A TISSUE- AND TIME-SPECIFIC  
 CC MANNER DURING DEVELOPMENT.  
 CC -!- SIMILARITY: INCLUDES 13.5 EGF-LIKE REPEATS AND 11 FIBRONECTIN  
 CC TYPE III-LIKE DOMAINS.  
 DR EMBL; M23121; GG TEN.  
 DR EMBL; X08031; GG TENAS1.  
 DR EMBL; X08030; GG TENAS8.  
 DR EMBL; J03641; GGCYTT1.  
 DR EMBL; M20816; GGCYTT2.  
 DR PIR; A30903; A30903.  
 DR PIR; A31930; A31930.  
 DR PIR; A33379; A33379.  
 DR PIR; B33379; B33379.  
 DR PIR; C33379; C33379.  
 DR PIR; S01292; S01292.  
 DR PROSITE; PS00022; EGF.  
 KW GLYCOPROTEIN; CELL ADHESION; TANDEM REPEAT; EGF-LIKE DOMAIN;  
 KW EXTRACELLULAR MATRIX; SIGNAL; ALTERNATIVE SPLICING.  
 FT SIGNAL 1 22  
 FT PROPEP 23 33  
 FT CHAIN 34 1808 TENASCIN.  
 FT DOMAIN 119 147 4 HEPTAD REPEATS (PROBABLE COILED COIL).  
 FT DOMAIN 176 590 13.5 EGF-TYPE REPEATS.  
 FT REPEAT 176 187 EGF-LIKE 0 (PARTIAL).  
 FT REPEAT 187 218 EGF-LIKE 1.  
 FT REPEAT 219 249 EGF-LIKE 2.  
 FT REPEAT 250 280 EGF-LIKE 3.  
 FT REPEAT 281 311 EGF-LIKE 4.  
 FT REPEAT 312 342 EGF-LIKE 5.  
 FT REPEAT 343 373 EGF-LIKE 6.  
 FT REPEAT 374 404 EGF-LIKE 7.  
 FT REPEAT 405 435 EGF-LIKE 8.  
 FT REPEAT 436 466 EGF-LIKE 9.  
 FT REPEAT 467 497 EGF-LIKE 10.  
 FT REPEAT 498 528 EGF-LIKE 11.  
 FT REPEAT 529 559 EGF-LIKE 12.  
 FT REPEAT 560 590 EGF-LIKE 13.  
 FT DOMAIN 591 680 FIBRONECTIN TYPE-III 1.  
 FT DOMAIN 681 771 FIBRONECTIN TYPE-III 2.  
 FT DOMAIN 772 862 FIBRONECTIN TYPE-III 3.  
 FT DOMAIN 863 954 FIBRONECTIN TYPE-III 4.  
 FT DOMAIN 955 1042 FIBRONECTIN TYPE-III 5.  
 FT DOMAIN 1043 1133 FIBRONECTIN TYPE-III 6.  
 FT DOMAIN 1134 1224 FIBRONECTIN TYPE-III 7.  
 FT DOMAIN 1225 1315 FIBRONECTIN TYPE-III 8.  
 FT DOMAIN 1316 1404 FIBRONECTIN TYPE-III 9.  
 FT DOMAIN 1405 1492 FIBRONECTIN TYPE-III 10.  
 FT DOMAIN 1493 1580 FIBRONECTIN TYPE-III 11.  
 FT SIMILAR 1589 1808 TO THE GLOBULAR DOMAIN OF THE BETA- AND  
 FT GAMMA-CHAINS OF FIBRINOGEN.  
 FT VARSPLIC 1043 1224 MISSING (IN 200 KD FORM).  
 FT VARSPLIC 1043 1315 MISSING (IN 190 KD FORM).

PVSARVATYLPAPE  
680



> 0 <  
0| 0 IntelliGenetics  
> 0 <

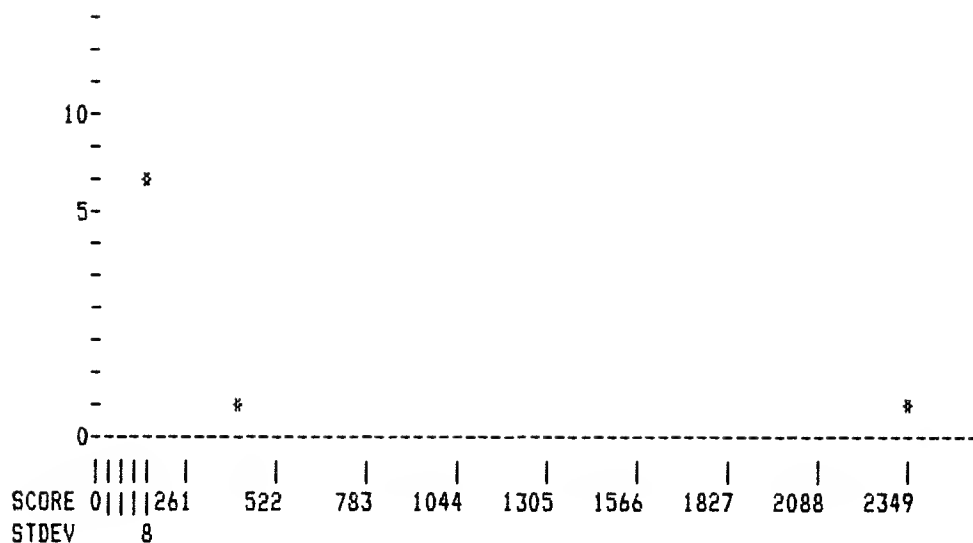
FastDB - Fast Pairwise Comparison of Sequences  
Release 5.4

Results file ellis-012-fig2ab.res made by shears on Tue 14 Sep 93 18:08:07-PDT.

Query sequence being compared: ELLIS-012-FIG2AB.SEQ (1-2350)  
Number of sequences searched: 144007  
Number of scores above cutoff: 3862

Results of the initial comparison of ELLIS-012-FIG2AB.SEQ (1-2350) with:  
Data bank : EMBL-NEW 7, all entries  
Data bank : GenBank 77, all entries  
Data bank : GenBank-NEW 6, all entries  
Data bank : UEMBL 35\_77, all entries

100000\*  
-  
N -\*  
U50000-  
M -  
B -  
E -  
R -  
-  
D -  
F10000-  
-  
S -  
E 5000-  
Q -  
U -  
E - \*  
N -  
C -  
E -  
S 1000-  
-  
-  
500-  
-  
-  
-  
-  
-  
- \*  
100-  
-  
-  
50-  
-  
-  
-



# PARAMETERS

Similarity matrix	Unitary	K-tuple	4
Mismatch penalty	1	Joining penalty	30
Gap penalty	1.00	Window size	32
Gap size penalty	0.33		
Cutoff score	10		
Randomization group	0		
Initial scores to save	40	Alignments to save	15
Optimized scores to save	0	Display context	50

# SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	36	36	14.60

Times:	CPU	Total Elapsed
	02:07:00.01	03:17:05.00

Number of residues:	169341811
Number of sequences searched:	144007
Number of scores above cutoff:	3862

Cut-off raised to 30.  
 Cut-off raised to 35.  
 Cut-off raised to 39.  
 Cut-off raised to 43.  
 Cut-off raised to 47.  
 Cut-off raised to 50.  
 Cut-off raised to 53.  
 Cut-off raised to 56.  
 Cut-off raised to 59.  
 Cut-off raised to 62.  
 Cut-off raised to 65.  
 Cut-off raised to 68.

The scores below are sorted by initial score.  
 Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Score	Score	Sig	Score
			Init.	Opt.		

**** 158 standard deviations above mean ****						
1. MUSTC41BB	Mouse T-cell receptor 4-1BB p	2350	2349	2349	158.43	0
**** 25 standard deviations above mean ****						
2. HUMILAX	Human activation dependent T	1419	412	806	25.75	0
**** 8 standard deviations above mean ****						
3. CBRR5A	Caenorhabditis briggsae DNA f	944	163	406	8.70	0
4. CLS88D0	Hamster EcoRI donor DNA fragm	3906	162	1019	8.63	0
5. XELAEIP	X.laevis amidating enzyme (AE	2733	157	960	8.29	0
**** 7 standard deviations above mean ****						
6. HUMUT5094	Human chromosome 4 STS UT5094	468	152	210	7.95	0
7. S53907	XRAR alpha 2=retinoic acid re	3240	151	961	7.88	0
8. PFASXC	Plasmodium falciparum sexual	2306	150	808	7.81	0
9. ACLRGNAL	A.laidlawii 16S ribosomal RNA	1508	146	568	7.53	0
10. HUMBIND	Human binding protein mRNA, p	3523	145	772	7.47	0
11. HSBIND	Human binding protein mRNA, p	3523	145	772	7.47	0
12. HSHB15RNA	Homo sapiens mRNA for HB15	1761	143	697	7.33	0
13. S53354	B-cell activation protein=B-G	2574	143	897	7.33	0
14. HSIL05	Human interleukin-2 (IL-2) ge	6684	142	737	7.26	0
15. RATTGFB	Rat transforming growth facto	6244	141	993	7.19	0
**** 6 standard deviations above mean ****						
16. PFAHRKPM	P.cynomolgi DNA homologous to	1563	136	636	6.85	0
17. HUMPALF1	Human mutant prealbumin gene	1913	136	798	6.85	0
18. HUMANYLOID	Homo sapiens amyloid protein	3725	136	757	6.85	0
19. HUMPALD	Human prealbumin gene, comple	7616	136	944	6.85	0
20. AMVCP	Arabis mosaic virus RNA-2, 3'	2406	135	918	6.78	0
21. HUMPALC	Human serum prealbumin gene.	7619	135	945	6.78	0
22. SCCHRIII	S.cerevisiae chromosome III c	315338	133	975	6.64	0
23. DROFATFA	Fruitfly fat facets mRNA.	8473	131	996	6.51	0
24. DROFATFB	Fruitfly fat facets mRNA.	8891	131	996	6.51	0
25. PIGFSHB	Pig follicle stimulating horm	929	130	398	6.44	0
26. ATCRPG	A.thaliana genes encoding gly	9619	130	962	6.44	0
27. NMUPA	M.musculus upstream region of	4431	129	879	6.37	0
28. MMGCSF	Mouse granulocyte colony-stim	1363	127	587	6.23	0
29. OCPMA1	O.cuniculus PMCA1 gene for pl	4479	126	800	6.16	0
30. S56304S1	AADC=aromatic L-amino acid de	1314	125	550	6.10	0
31. STAPT48CG	Plasmid pT48 (from S.aureus)	2475	125	713	6.10	0
32. RATOLFPROL	Rat olfactory protein mRNA, c	984	124	394	6.03	0
33. MMUPAACT	Mouse gene for urokinase plas	986	124	426	6.03	0
34. HUMHTF4	Human helix-loop-helix protei	2942	124	897	6.03	0
35. MUSFABP1	Mouse Fabpi gene, exons 1-4.	5039	124	843	6.03	0
**** 5 standard deviations above mean ****						
36. CEHER1GNA	C.elegans her-1 gene	6932	123	972	5.96	0
37. YSCMTAT92	yeast (s.cerevisiae) mitochon	365	121	179	5.82	0
38. M75767	CEL02A3S2 Caenorhabditis eleg	388	121	164	5.82	0
39. SCSPP91A	S.cerevisiae SPP91 gene	1665	121	698	5.82	0
40. YSCPRP21A	Saccharomyces cerevisiae nucl	2180	121	709	5.82	0

# 1. ELLIS-012-FIG2AB.SEQ (1-2350)

MUSTC41BB Mouse T-cell receptor 4-1BB protein mRNA, complete

LOCUS MUSTC41BB 2350 bp ss-mRNA ROD 15-SEP-1989  
DEFINITION Mouse T-cell receptor 4-1BB protein mRNA, complete cds.  
ACCESSION J04492  
KEYWORDS T-cell receptor.  
SOURCE Mouse (strain C57BL/6) T-lymphocyte cell lines L2 and L3, cDNA to mRNA.  
ORGANISM Mus musculus  
Eukaryota; Animalia; Chordata; Vertebrata; Mammalia; Theria;  
Eutheria; Rodentia; Myomorpha; Muridae; Murinae.  
REFERENCE 1 (bases 1 to 2350)  
AUTHORS Kwon,B.S. and Weissman,S.M.  
TITLE cDNA sequences of two inducible T-cell genes  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 86, 1963-1967 (1989)  
STANDARD full automatic

COMMENT Draft entry and clean copy of sequence for [1] kindly provided by  
B.S.Kwon, 17-MAR-1989.

FEATURES Location/Qualifiers  
sig\_peptide 146..214  
/codon\_start=1  
/note="4-1BB protein signal peptide"  
mat\_peptide 215..913  
/codon\_start=1  
/note="4-1BB protein"  
CDS 146..916  
/note="4-1BB protein precursor"  
/codon\_start=1  
/translation="MGNNCYNVVVIVLLLVGCEKVGAVQNSCDNCQPGTFCRKYNPVC  
KSCPPSTFSSIGGQPNICRVACAGYFRFKKCSSTHNAECECIEGFHCLGPQCTRCE  
KDCRPGQELTKQGCKTCSLGTFNQNGTGVCRPWTNCSLDGRSVLKTGTTEKDVVCGP  
PVVSFSPSTTISVTEGGPGGHSLOVLTFLALTSALLLALIFITLLFSVLKWKIRKKF  
PHIFKQPFKKTGAQQEEDACSCRCPEEEGGGGGYEL"

BASE COUNT 590 a 561 c 589 g 607 t 3 others  
ORIGIN Unreported.

Initial Score = 2349 Optimized Score = 2349 Significance = 158.43  
Residue Identity = 99% Matches = 2349 Mismatches = 1  
Gaps = 0 Conservative Substitutions = 0

X 10 20 30 40 50 60 70  
ATGTCCATGAAGTCTGAGTGGATAAACAGCACGGGATATCTCTGTCTAAAGGAATATTACTACACCAGGAA  
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
ATGTCCATGAAGTCTGAGTGGATAAACAGCACGGGATATCTCTGTCTAAAGGAATATTACTACACCAGGAA  
X 10 20 30 40 50 60 70

80 90 100 110 120 130 140  
AAGGACACATTCGACAACAGGAAAGGAGCCTGTACAGAAAACACAGTGTCTGTGCATGTGACATTTTCGC  
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
AAGGACACATTCGACAACAGGAAAGGAGCCTGTACAGAAAACACAGTGTCTGTGCATGTGACATTTTCGC  
80 90 100 110 120 130 140

150 160 170 180 190 200 210  
CATGGGAAACAACTGTTACAACGTGGTGGTATTGTGCTGCTAGTGGGCTGTGAGAAGGTGGGAGCCGT  
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
CATGGGAAACAACTGTTACAACGTGGTGGTATTGTGCTGCTAGTGGGCTGTGAGAAGGTGGGAGCCGT  
150 160 170 180 190 200 210

220 230 240 250 260 270 280  
GCAGAACTCCTGTGATAACTGTGAGCCTGGTACTTTCTGCAGAAAATACAATCCAGTCTGCAAGAGCTGCCC  
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
GCAGAACTCCTGTGATAACTGTGAGCCTGGTACTTTCTGCAGAAAATACAATCCAGTCTGCAAGAGCTGCCC  
220 230 240 250 260 270 280

290 300 310 320 330 340 350 360  
TCCAAGTACCTTCTCCAGCATAGGTGGACAGCCGAAGTGAACATCTGCAGAGTGTGTGCAGGCTATTTTCA  
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
TCCAAGTACCTTCTCCAGCATAGGTGGACAGCCGAAGTGAACATCTGCAGAGTGTGTGCAGGCTATTTTCA  
290 300 310 320 330 340 350 360

370 380 390 400 410 420 430  
GTTCAAGAAGTTTGTCTCTACCCACAACGCGGAGTGTGAGTGCATTGAAGGATTCCATTGCTTGGGGCC  
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
GTTCAAGAAGTTTGTCTCTACCCACAACGCGGAGTGTGAGTGCATTGAAGGATTCCATTGCTTGGGGCC  
370 380 390 400 410 420 430

440 450 460 470 480 490 500  
ACAGTGCACCAGATGTGAAAAGGACTGCAGGCTGGCCAGGAGCTAACGAAGCAGGTTGCAAAACCTGTAG  
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
ACAGTGCACCAGATGTGAAAAGGACTGCAGGCTGGCCAGGAGCTAACGAAGCAGGTTGCAAAACCTGTAG  
440 450 460 470 480 490 500

510 520 530 540 550 560 570  
CTTGGGAACATTTAATGACCAGAACGGTACTGGCGTCTGTGACCCCTGGACGAACTGCTCTCTAGACGGAAG  
|||||  
CTTGGGAACATTTAATGACCAGAACGGTACTGGCGTCTGTGACCCCTGGACGAACTGCTCTCTAGACGGAAG  
510 520 530 540 550 560 570

580 590 600 610 620 630 640  
GTCTGTGCTTAAGACCGGGACCACGGAGAAGGACGTGGTGTGTGGACCCCTGTGGTGAGCTTCTCTCCCAG  
|||||  
GTCTGTGCTTAAGACCGGGACCACGGAGAAGGACGTGGTGTGTGGACCCCTGTGGTGAGCTTCTCTCCCAG  
580 590 600 610 620 630 640

650 660 670 680 690 700 710 720  
TACCACCATTTCTGTGACTCCAGAGGGAGGACCAGGAGGGGACTCCTTGCAGGTCCTTACCTTGTTCCTGGC  
|||||  
TACCACCATTTCTGTGACTCCAGAGGGAGGACCAGGAGGGGACTCCTTGCAGGTCCTTACCTTGTTCCTGGC  
650 660 670 680 690 700 710 720

730 740 750 760 770 780 790  
GCTGACATCGGCTTTGCTGCTGGCCCTGATCTTCATTACTCTCCTGTTCTCTGTGCTCAAATGGATCAGGAA  
|||||  
GCTGACATCGGCTTTGCTGCTGGCCCTGATCTTCATTACTCTCCTGTTCTCTGTGCTCAAATGGATCAGGAA  
730 740 750 760 770 780 790

800 810 820 830 840 850 860  
AAAATTCCCCACATATTCAAGCAACCATTTAAGAAGACCACTGGAGCAGCTCAAGAGGAAGATGCTTGTAG  
|||||  
AAAATTCCCCACATATTCAAGCAACCATTTAAGAAGACCACTGGAGCAGCTCAAGAGGAAGATGCTTGTAG  
800 810 820 830 840 850 860

870 880 890 900 910 920 930  
CTGCCGATGTCCACAGGAAGAAGAAGGAGGAGGAGGCTATGAGCTGTGATGTACTATCCTAGGAGATGT  
|||||  
CTGCCGATGTCCACAGGAAGAAGAAGGAGGAGGAGGCTATGAGCTGTGATGTACTATCCTAGGAGATGT  
870 880 890 900 910 920 930

940 950 960 970 980 990 1000  
GTGGGCCGAAACCGAGAAGCACTAGGACCCACCATCCTGTGGAACAGCACAAAGCAACCCACCACCCTGTT  
|||||  
GTGGGCCGAAACCGAGAAGCACTAGGACCCACCATCCTGTGGAACAGCACAAAGCAACCCACCACCCTGTT  
940 950 960 970 980 990 1000

1010 1020 1030 1040 1050 1060 1070 1080  
CTTACACATCATCCTAGATGATGTGTGGGCGGCACCTCATCCAAGTCTCTTCTAACGCTAACATATTTGTC  
|||||  
CTTACACATCATCCTAGATGATGTGTGGGCGGCACCTCATCCAAGTCTCTTCTAACGCTAACATATTTGTC  
1010 1020 1030 1040 1050 1060 1070 1080

1090 1100 1110 1120 1130 1140 1150  
TTTACCTTTTTAAATCTTTTTTAAATTTAAATTTATGTGTGTGAGTGTTCCTGCCTGTATGCACAC  
|||||  
TTTACCTTTTTAAATCTTTTTTAAATTTAAATTTATGTGTGTGAGTGTTCCTGCCTGTATGCACAC  
1090 1100 1110 1120 1130 1140 1150

1160 1170 1180 1190 1200 1210 1220  
GTGTGTGTGTGTGTGTGTGTGACACTCCTGATGCCTGAGGAGGTGAGAAGAGAAAGGGTTGGTTCCATAAGA  
|||||  
GTGTGTGTGTGTGTGTGTGTGACACTCCTGATGCCTGAGGAGGTGAGAAGAGAAAGGGTTGGTTCCATAAGA  
1160 1170 1180 1190 1200 1210 1220

1230 1240 1250 1260 1270 1280 1290  
ACTGGAGTTATGGATGGCTGTGAGCCGGNNGATAGGTGCGGACGGAGACCTGTCTTCTTATTTAACGTGA  
|||||  
ACTGGAGTTATGGATGGCTGTGAGCCGGNNGATAGGTGCGGACGGAGACCTGTCTTCTTATTTAACGTGA  
1230 1240 1250 1260 1270 1280 1290

1300 1310 1320 1330 1340 1350 1360  
CTGTATAATAAAAAAAAAAATGATATTTCCGGAATTGTAGAGATTGTCTGACACCCTTCTAGTTAATGATCT  
|||||  
CTGTATAATAAAAAAAAAAATGATATTTCCGGAATTGTAGAGATTCTCTGACACCCTTCTAGTTAATGATCT  
1300 1310 1320 1330 1340 1350 1360

1370 1380 1390 1400 1410 1420 1430 1440  
AAGAGGAATTGTTGATACGTAGTATACTGTATATGTGTATGTATATGTATATATAAGACTCTTTTA  
|||||  
AAGAGGAATTGTTGATACGTAGTATACTGTATATGTGTATGTATATGTATATATAAGACTCTTTTA  
1370 1380 1390 1400 1410 1420 1430 1440

1450 1460 1470 1480 1490 1500 1510  
CTGTCAAAGTCAACCTAGAGTGTCTGTTTACCAGGTCAATTTTATTGGACATTTTACGTACACACACACAC  
|||||  
CTGTCAAAGTCAACCTAGAGTGTCTGTTTACCAGGTCAATTTTATTGGACATTTTACGTACACACACACAC  
1450 1460 1470 1480 1490 1500 1510

1520 1530 1540 1550 1560 1570 1580  
ACACACACACACACGTTTATACTACGTACTGTTATCGGTATTCTACGTCATATAATGGGATAGGGTAAAA  
|||||  
ACACACACACACACGTTTATACTACGTACTGTTATCGGTATTCTACGTCATATAATGGGATAGGGTAAAA  
1520 1530 1540 1550 1560 1570 1580

1590 1600 1610 1620 1630 1640 1650  
GGAAACCAAAGAGTGAGTGATATTATTGTGGAGGTGACAGACTACCCCTTCTGGGTACGTAGGGACAGACCT  
|||||  
GGAAACCAAAGAGTGAGTGATATTATTGTGGAGGTGACAGACTACCCCTTCTGGGTACGTAGGGACAGACCT  
1590 1600 1610 1620 1630 1640 1650

1660 1670 1680 1690 1700 1710 1720  
CCTTCGGACTGTCTAAAACTCCCCTTAGAAGTCTCGTCAAGTTCCCGGACGAAGAGGACAGAGGAGACACAG  
|||||  
CCTTCGGACTGTCTAAAACTCCCCTTAGAAGTCTCGTCAAGTTCCCGGACGAAGAGGACAGAGGAGACACAG  
1660 1670 1680 1690 1700 1710 1720

1730 1740 1750 1760 1770 1780 1790 1800  
TCCGAAAAGTTATTTTCCGGCAAATCCTTTCCCTGTTTCGTGACACTCCACCCCTTGTGGACACTTGAGTG  
|||||  
TCCGAAAAGTTATTTTCCGGCAAATCCTTTCCCTGTTTCGTGACACTCCACCCCTTGTGGACACTTGAGTG  
1730 1740 1750 1760 1770 1780 1790 1800

1810 1820 1830 1840 1850 1860 1870  
TCATCCTTGCGCCGGAAGGTCAGGTGGTACCCGTCTGTAGGGGCGGGGAGACAGAGCCGCGGGGAGCTACG  
|||||  
TCATCCTTGCGCCGGAAGGTCAGGTGGTACCCGTCTGTAGGGGCGGGGAGACAGAGCCGCGGGGAGCTACG  
1810 1820 1830 1840 1850 1860 1870

1880 1890 1900 1910 1920 1930 1940  
AGAATCGACTCACAGGGCGCCCCGGGCTTCGCAAATGAAACTTTTTTAATCTCACAAGTTTCGTCCGGGCTC  
|||||  
AGAATCGACTCACAGGGCGCCCCGGGCTTCGCAAATGAAACTTTTTTAATCTCACAAGTTTCGTCCGGGCTC  
1880 1890 1900 1910 1920 1930 1940

1950 1960 1970 1980 1990 2000 2010  
GGCGGACCTATGGCGTCGATCCTTATTACCTTATCCTGGCGCCAAGATAAAACAACCAAAAGCCTTGACTCC  
|||||  
GGCGGACCTATGGCGTCGATCCTTATTACCTTATCCTGGCGCCAAGATAAAACAACCAAAAGCCTTGACTCC  
1950 1960 1970 1980 1990 2000 2010

2020 2030 2040 2050 2060 2070 2080  
GGTACTAATTCTCCCTGCCGGCCCCGTAAGCATAACGCGGCGATCTCCACTTTAAGAACCCTGGCCGCGTTC  
|||||  
GGTACTAATTCTCCCTGCCGGCCCCGTAAGCATAACGCGGCGATCTCCACTTTAAGAACCCTGGCCGCGTTC  
2020 2030 2040 2050 2060 2070 2080

```

2090      2100      2110      2120      2130      2140      2150      2160
TGCCTGGTCTCGCTTTGTAACGGTTCTTACAAAAGTAATTAGTTCTTGCTTTACAGCCTCCAAGCTTCTGC
|||||
TGCCTGGTCTCGCTTTGTAACGGTTCTTACAAAAGTAATTAGTTCTTGCTTTACAGCCTCCAAGCTTCTGC
2090      2100      2110      2120      2130      2140      2150      2160

      2170      2180      2190      2200      2210      2220      2230
TAGTCTATGGCAGCATCAAGGCTGGTATTTGCTACGGCTGACCGCTACGCCGCCGCAATAAGGGTACTGGGC
|||||
TAGTCTATGGCAGCATCAAGGCTGGTATTTGCTACGGCTGACCGCTACGCCGCCGCAATAAGGGTACTGGGC
      2170      2180      2190      2200      2210      2220      2230

      2240      2250      2260      2270      2280      2290      2300
GGCCCGTCGAAGGCCCTTTGGTTTCAGAAACCCAAGGCCCCCTCATACCAACGTTTCGACTTTGATTCTTG
|||||
GGCCCGTCGAAGGCCCTTTGGTTTCAGAAACCCAAGGCCCCCTCATACCAACGTTTCGACTTTGATTCTTG
      2240      2250      2260      2270      2280      2290      2300

      2310      2320      2330      2340      X
CCGGTACGTGGTGGTGGGTGCCTTAGCTCTTCTCGATAGTTAGAC
|||||
CCGGTACGTGGTGGTGGGTGCCTTAGCTCTTCTCGATAGTTAGAC
      2310      2320      2330      2340      2350

```

## 2. ELLIS-012-FIG2AB.SEQ (1-2350)

HUMILAX Human activation dependent T cell mRNA, complete c

```

LOCUS      HUMILAX      1419 bp ss-mRNA      PRI      30-APR-1993
DEFINITION Human activation dependent T cell mRNA, complete cds.
ACCESSION  L12964
KEYWORDS   cell surface receptor; nerve growth factor receptor;
           tumor necrosis factor receptor.
SOURCE      Homo sapiens cDNA to mRNA.
ORGANISM    Homo sapiens
            Eukaryota; Animalia; Chordata; Vertebrata; Mammalia; Theria;
            Eutheria; Primates; Haplorhini; Catarrhini; Hominidae.
REFERENCE   1 (bases 1 to 1419)
AUTHORS     Schwarz,H., Tuckwell,J.E. and Lotz,M.
TITLE       Nucleotide sequence of ILA, a cDNA encoding a new member of the
           human nerve growth factor/tumor necrosis factor receptor family
JOURNAL      Unpublished (1993)
STANDARD    full automatic
FEATURES     Location/Qualifiers
             5'UTR      1..139
             3'UTR      908..1419
             polyA_signal 1369..1374
             polyA_site  1419
             source      1..1419
                        /organism="Homo sapiens"
                        /cell_type="transformed T lymphocyte /cell_line SLB-1"
                        /sequenced_mol="cDNA to mRNA"
             CDS         140..907
                        /gene="ILA"
                        /note="ILA= induced by lymphocyte activation"
                        /codon_start=1
                        /translation="MGNSCYNIVATLLLVLFNFERTRSLQDPCSNCPAGTFCDNNRNQI
                        CSPCPPNSFSSAGGQRTCDICRQCKGVFRTRKECSSTSNAECDCTPGFHCLGAGCSMC
                        EQDCRQGEELTKKGCKDCCFGTNDQKRGICRPWTNCSLDGKSVLVNGTKERDVVCGP
                        SPADLSPGASSVTPPAPAREPGHSPQIISFFLALTSTALLFLLFFLTLRFSVVKRGK
                        KLLYIFKQPFMRPVQTTQEEEDGCSCRFEEEEGGCEL"
BASE COUNT   373 a   340 c   342 g   364 t
ORIGIN

```

Initial Score = 412 Optimized Score = 806 Significance = 25.75

Residue Identity = 40% Matches = 904 Mismatches = 477

[illegible]



700 710 720 730 740 750 760  
770 780 790 800 810 820 830  
CCTGTTCTCTGTGCTCAAATGGATCAGGAAAAAATCCCCACATATTCAAGCAACCATTTAAGAAGACCAC  
|| ||||| | ||| || ||| ||| ||| ||||| ||||| ||||| |||||  
CCGTTTCTCTGTTGTTAAACGGGGCAGAAAGAACTCCTGTATATATTCAAACAACCATTATG-AGACCAG  
770 780 790 800 810 820 830  
840 850 860 870 880 890 900  
T-GGAGCAGCTCAAGAGGAAGATGCTGTAGCTGCCGATGTCCACAGGAAGAAGAAGGAGGAGGAGGAGGCT  
| | | ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
TACAACTACTCAAGAGGAAGATGGCTGTAGCTGCCGATTTCAGAGAAGAAGAAGGAGGATGTGAA--CT  
840 850 860 870 880 890 900  
910 920 930 940 950 960 970  
ATGAGCTGTGATGTACTATCCTAGGAGATGTGTGGGCCGAAACCGAGAAGCA-CTAGGA---CCCCACCATC  
||| ||| ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
GTGAAATG-GAAGT-CAA---TAGG-GCTGT-TGGGACTTTCTTGAAAAGAAGCAAGGAAATATGAGTCATC  
910 920 930 940 950 960  
980 990 1000 1010 1020 1030 1040  
CTGTGGAACAG----CACAAGCAACCCACCACCCTGTTCTTACACATCAT-CCTAGATGATGTGTGGGCGC  
| | |||| | ||||| ||||| ||| ||||| ||| ||| ||| |||  
CGCTATCACAGCTTTCAAAGCAAGAACACCATCCT-----ACATAATACCAG--GAT-----TCCC  
970 980 990 1000 1010 1020  
1050 1060 1070 1080 1090 1100  
GCACCTCATCCAAGTCTCTTCT-AACGCTAACATATTTGTCTTT-----ACCTTTTTTAAATCTTTTT  
|| | || ||||| || ||| || ||||| ||||| ||||| ||||| ||||| |||||  
CCAACACA---CGTCTTTTTCTAAATGCCAATGAGTTGGCCTTTAAAAATGCACCACTTTTTTTTTTTTTT  
1030 1040 1050 1060 1070 1080 1090  
1110 1120 1130 1140 1150  
TAA----ATTTAAATTT-----TATG-TG-TGTG-AGT-GTTTTGCC-TGCCTGTATGCACACGTG-TGTG  
| | | ||| | ||||| ||||| ||||| ||||| ||||| ||||| |||||  
GGACAGGGTCTCACTCTGTCACCAGGCTGGAGTGCAGTGGCACCACCATGGCTCTCTGCAGCCTTGACCTC  
1100 1110 1120 1130 1140 1150 1160  
1160 1170 1180 1190 1200 1210 1220  
TGTGTGTGTGTGTGACACTCCT-----GATGCCTGAGGAGGTGAGAAGAGAAAGGGTTGGTTCCATAAGAA  
|| | || ||||| || ||||| || ||||| ||||| ||||| ||||| |||||  
TGGGAGCTCAAGTGATCCTCCTGCCTCAGTCTCCTGAGTAGCT-GGAACTACAAGGAAGGG--CCACCACAC  
1170 1180 1190 1200 1210 1220 1230  
1230 1240 1250 1260 1270 1280 1290  
CTGGAGTTA-TGGATGGCTGTGAGCCGNNNGATAGGTGGGACGGAGACCTGTCTT-----CTTATTTTAA  
|| | ||| | ||||| || ||||| || ||||| || ||||| || |||||  
CT-GACTAACTTTTTTGTTTTTTGTGG---TAAAGAT-GGCATTTGCCATGTTGTACAGGCTGGTCTCAA  
1240 1250 1260 1270 1280 1290 1300  
1300 1310 1320 1330 1340 1350  
CGTGACT-GTATAATAAAAAAAAT--GATATTTGGGAATTGTAGAGATTGTCCTGACA--CCCTTCTAG  
||| || ||| | ||||| ||||| ||||| ||||| ||||| |||||  
----ACTCCTAGGTTCACTTTGGCCTCCCAAAGTGCTGGGATTACAGACA-TGAACTGCCAGGCCCGGCCA-  
1310 1320 1330 1340 1350 1360  
1360 1370 1380 1390 1400 1410 1420 1430  
TTAATGATCTAAGAGGAATTGTTGATACGTAGTATACTGTATATGTGTATGTATATGTATATGTATATATAA  
||| || | ||||| || ||||| || ||||| || ||||| || |||||  
-AAATAAT---GCACCACCTT-TTAACA-GAA-CAGAC---AGATGAGGACAGAGCTGGTGAT  
1370 1380 1390 1400 1410 X  
1440 1450 1460 1470  
GACTCTTTTACTGTCAAAGTCAACCTAGAGTGTCTGTTA

CBRR5A *Caenorhabditis briggsae* DNA for 5S ribosomal RNA (

1570	1540	1550	1540	1570	1580	1590
------	------	------	------	------	------	------



LOCUS CL588D0 3906 bp DNA ROD 20-MAY-1992  
 DEFINITION Hamster EcoRI donor DNA fragment for S88 aprt insertion  
 ACCESSION X14996 X13999 X14000  
 KEYWORDS Alu repetitive sequence; insertion sequence; repetitive sequence.  
 SOURCE chinese hamster  
 ORGANISM Cricetulus longicaudatus  
 Eukaryota; Animalia; Metazoa; Chordata; Vertebrata; Mammalia;  
 Theria; Eutheria; Rodentia; Myomorpha; Cricetidae; Cricetinae;  
 Cricetini.  
 REFERENCE 1 (bases 361 to 690; 1481 to 2640)  
 AUTHORS Nalbantoglu,J., Miles,C. and Neuth,M.  
 TITLE Insertion of Unique and Repetitive DNA Fragments into the aprt  
 Locus of Hamster Cells  
 JOURNAL J. Mol. Biol. 200, 449-459 (1988)  
 STANDARD full automatic  
 REFERENCE 2 (bases 1 to 3906)  
 AUTHORS Neuth,M.  
 TITLE Direct Submission  
 JOURNAL Submitted (13-FEB-1989) to the EMBL Data Library.  
 STANDARD full automatic  
 COMMENT \*source: clone\_library=lambda NM1149.;

See x07513 for S88 mutant seq.

FEATURES Location/Qualifiers  
 misc\_feature 572..624  
 /note="CHD Alu repeat"  
 conflict 1625  
 /citation=[1]  
 /note="c is g in [1]"  
 misc\_feature 1734..2018  
 /note="S88 donor DNA"

BASE COUNT 1048 a 945 c 905 g 1008 t  
 ORIGIN

Initial Score = 162 Optimized Score = 1019 Significance = 8.63  
 Residue Identity = 49% Matches = 1238 Mismatches = 954  
 Gaps = 325 Conservative Substitutions = 0

```

                                     X      10
                                     ATGTCCATGA----ACTGC--T
                                     | | | | | | | | | | | |
TCAGGGTTCACCATATCCTGAAGAGCAGATGAGACACCAGGCTTTTACCTTCCATCAGGAGACAACCAT
  940      950      960      970      980 X      990      1000

      20      30      40      50      60      70      80
GAGTGGG-TAAACAGCACGGGATATCTCTGTCTAAAGGAAT-ATTACTACACCAGGAAAAGGACA-CATTGG
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
GATTGGACTAGCCAG-ACCGCA---GCCTGTGAGCCGTTCTAATCAATCCCTACAAATAGAACATTTTCT
  1010     1020     1030     1040     1050     1060     1070

      90      100      110      120      130      140      150
AC-AACAGGAAAGGAGC-CTGT----CACAGAAAACCACAGTGTCTGTGCATGTGACATTTGCCATGGGA
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
TCCATCAGTTCTGTTCGGTTGGAGAACCCAGACTAATACAATCTCC---CATGTTCCAGTT--TCATGTGT
  1080     1090     1100     1110     1120     1130

      160      170      180      190      200      210
AACAAC--TGTTACAACGTGGT--GGTCATTGTGCTGCTAGTGGGCTGTGAG-AAGGTGGGAGCCGTGC
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
CACTACCATGTGACAGGCCAGTAGGGTGAGTGGGCT-CAGC---TGGGCTCTGCCCCAGGTGCGGATCAAAC
  1140     1150     1160     1170     1180     1190     1200

      220      230      240      250      260      270      280
AGA--ACTC--CTGTGATAA-CTGTGAGCCTGGTACTTTCTGCAGAA---AATACAATCC-AGTCTGCA-AG
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
AGACAGCTCTCTTTCGCAAAAGCTCTCATCTCAGCTTCTCTC-AGAACCCAGCTCTATCCCAATATCCATAC

```

1210 1220 1230 1240 1250 1260 1270  
290 300 310 320 330 340 350  
AGCTGCCCTCCAAGTACCTTCTCCAGCATAGGTGGACAGCCGAAGTGAACATCTGCAGAGTGTGTGCAGGC  
| | | | | | | | | | | | | | | | | | | | | |  
A----CCCCACTTGCAGC-TGTAGGACTGGGGTCCCTTTCCAGCTG-GATGTCAGC-GAGGGT---CA-CC  
1280 1290 1300 1310 1320 1330  
360 370 380 390 400 410  
TATTTCAAGTTCAAGAAGTTTGTCTCTACCCACAACGCGGAGTGTGAGTGCATTGAA-----GGA  
| | | | | | | | | | | | | | | | | | | | | |  
T-TAGCAACTGACAGACTTCTCATTCTTCATCTTCAAGGC--AGCAAGATCACCTCAAATCCTCCTCCTGA  
1340 1350 1360 1370 1380 1390 1400  
420 430 440 450 460 470 480  
TTCCA--TTGCTTGG-GGCCAC-AGTGCA-CCAG-ATGTGAAAAGGACTGCAGGCCTGGCCAGGAGCTAACG  
| | | | | | | | | | | | | | | | | | | | | |  
GTCAAACCTCCCTGGCTTCTCTTCTGCACCCAGCCAGTG-ACAGCTCT-CTGCTCT-GACAGGTTCAAGACC  
1410 1420 1430 1440 1450 1460 1470  
490 500 510 520 530 540 550  
AAGCAGGGTTG-CAAAAC-CTGTAGCTTGGGAACATTTAATGACCAGAACGGTACTG-GCGTCTGTCCAGCC  
| | | | | | | | | | | | | | | | | | | | | |  
CA-CAGGATGGTCTGTTCTCATTACCT--CCACCCTTCACGGATCACACAGGAAGTGAGGGAGGGACAACCTC  
1480 1490 1500 1510 1520 1530 1540  
560 570 580 590 600 610 620  
TGGACGAAGTCTCTCTAGACGGA-AGGTCTGTGCTTAAGACCGGGACCACGAGAAGGACGTGGTGTGTGG  
| | | | | | | | | | | | | | | | | | | | | |  
CAGA-TTCCCGTCCAGCACAGGGACAGGGAGGTGCACCAG-CTG--AAGACAGAGTAG--AGTTGAATGCTG  
1550 1560 1570 1580 1590 1600 1610  
630 640 650 660 670 680 690  
ACCCCTGTGGTGAGCTTCTCTCCAGTACCACCATTTCTGTGACTCCAGAGGGAGGACCAGGAGGGCACTC  
| | | | | | | | | | | | | | | | | | | | | |  
ACCAAGTCAGGACACTTTGAGGAGATGGCTGTTCAAGACTGAG-CCCAAATTGCG--CCTGTGAAGAACTG  
1620 1630 1640 1650 1660 1670  
700 710 720 730 740 750 760  
CTTGACAGTCTTACCT---TGTTCTGGC--GCTGACATCGGCTTTGCTGCTGGCCCTGATCTTCATTACT  
| | | | | | | | | | | | | | | | | | | | | |  
ATT-CAGATAC-TACATGAGTGTGATTGACAAGCAGGAAACAGCATGG--GCT----AAGATGTAAGCA--  
1680 1690 1700 1710 1720 1730 1740  
770 780 790 800 810 820 830  
CTCCTGTTCTCTGTGCTCAAATGGATCAGGAAAAAATTC-CCCCACATAT-TCAAGCAACCATTTAAGAAGA  
| | | | | | | | | | | | | | | | | | | | | |  
---GAGAAATCT-TG-ACACATACA-CA--CACTCATACAGCCCTGTTATCACAATAAGACATGTA--TTGG  
1750 1760 1770 1780 1790 1800  
840 850 860 870 880 890 900  
CCACTGGAGCAGCT-CAAGAGGAAGATGCTGTAGCTGCCGATGTCCACAGGAAGAAGAAGGAGGAGGAGGA  
| | | | | | | | | | | | | | | | | | | | | |  
TCACTCCCAAAGGTCTTGTGCCCCGTG-AGGTCCCTCC-CTCACC-CTGTCTGCTCCCCAGACCCAGGA  
1810 1820 1830 1840 1850 1860 1870  
910 920 930 940 950 960  
GGCTATGAGCTGTGATGTACTA-TCCTAG-GAGATGTGTGGGCCGAAACCGAGAAG-CACTAGGACCCCA--  
| | | | | | | | | | | | | | | | | | | | | |  
AACCATAA--TGTGCT-TTCTATTCTAGAGATATATTTTGG--GCTTTCTAGAAGTTTCTACAAACTCAGA  
1880 1890 1900 1910 1920 1930  
970 980 990 1000 1010 1020  
-CCA-TCCTGTGGAACAGCACAAGC-----AACCCACCAACCTGTTCTTACACATCATCCTAGATGA-  
| | | | | | | | | | | | | | | | | | | | | |  
GACACTGTTCTGATTAAACAAGTCTCTGTTAAACCCCAAGCCCTGCGCATCTCACA-----ACCTAAC



2700      2710      2720      2730      2740      2750  
 1770      1780      1790      1800      1810      1820      1830  
 CGTGACACTCCACCCCTTGTGGACACTTGAGTGTTCATCCTTGCGCCGGAAGGTCAGGTGGT---ACCCGT-C  
 |    | | | |    |    | |    | | | |    | |    | | | | | |    | | | |  
 C-AAACAATACAAAACTTAAAAAAAAGAGT-TAA--AAAGCTATGTATGCTAAGATACTCAAATCAGTAT  
 2760      2770      2780      2790      2800      2810      2820  
  
 1840      1850      1860      1870      1880      1890      1900  
 TGTAGGGGCGGGGAGACAGAGC--CGCGGGGGAGCTACGAGAATCGACTCACAGGGCGCCCGGGCTTCGCA  
 | | | | |    | | | |    | | | |    | | | |    | |    | |    | | | | | |  
 TGTATGCGTAAGTTGATAACCAAAAAACAAAACCATAACAAAGGACAGAAAGATGGCCTC-AGCT-GGTA  
 2830      2840      2850      2860      2870      2880      2890  
  
 1910      1920      1930      1940      1950      1960  
 AA-----TGAAACTTTTTAATCTCACAAGTTTCGT--CCGG-GCTCGGCGGACCTATGGCGTCGATCCTT  
 | |    | |    | |    | |    | | | |    | | | |    | |    | | | | | |  
 AAGGTGTGTGTACGGCCTTGA--TCGTGAGTTCCATCCCCGGAACCTATCTGGTGAAGGAGAGGAT--GG  
 2900      2910      2920      2930      2940      2950      2960  
  
 1970      1980      1990      2000      2010      2020      2030  
 ATTACCTTATCCTGGCGCAAGATAAAACAACCAAAAGC--CTTG-ACTCCGGTACTAATTCTCCCTGCCGG  
 | |    | |    | | | |    | |    | | | |    | | | |    | |    | | | | | |  
 ACTCTCTTGCTCTGTCCCCTA-CTTTCACAGGCACAGGCTTCCTGCACACAG--ACGCATAC-ACATAAATG  
 2970      2980      2990      3000      3010      3020      3030  
  
 2040      2050      2060      2070      2080      2090      2100  
 CCCCCTGTA--GCATAACGCGGCGA-TCTCCACTTTAAGAACCTGGCCGCTTCTGCCTGGTC-TCGCTTTC  
 |    | |    | | | |    | | | |    | | | | | |    | |    | | | | | |  
 TAACCTAAATTGTGTTAAGTTTCCACTGTCAAGTGAACAACCCAGCGAGTTTCTGTATAGTCATAGCCT--  
 3040      3050      3060      3070      3080      3090      3100  
  
 2110      2120      2130      2140      2150      2160      2170  
 GTAAACGGTTC--TTACAAAAGTAATTAGTTCTTGCTTTTACGCCTCCAAGCTTCTGCTAGTCTATGGCAGCA  
 | | |    |    | | | |    |    | | | |    | | |    | |    | |    | | | | | |  
 GTAGCCATCACCATTACCAAGTGCAGAAGATTTTTATCACA-CAAGAAAGGAGC-CCCATGCCATCGCATCA  
 3110      3120      3130      3140      3150      3160      3170  
  
 2180      2190      2200      2210      2220      2230      2240  
 TCAAG-GCTGGTATTTGCTACGGCTGACCGCTACGCCGCCCAATAAGGGTACTGGGCGGCCCGTCAAGGC  
 |    | |    | | | |    | | | |    | | | |    | |    | |    | | | | | |  
 CCCCAGCCCTGCAGCTGCT-CATCTG-CCG-TCCATCTCTG-AAT-----T--TG-----CCTTCTACAGA  
 3180      3190      3200      3210      3220  
  
 2250      2260      2270      2280      2290      2300      2310  
 CCTTTGGTTTCAGAAACCAAGGCCCCCTCATACCAACGTTT-CGACTTTGATTCTTGCCGGTACGTGGT-  
 |    | | | |    | | | |    | | | |    | | | |    | | | |    | | | | | |  
 CTCTTCATAT-AAAGAGTCAA--TCAACCT----CCAGCCTTTGCGTCTGGCGTATTTCCCTGAGCGCAGTA  
 3230      3240      3250      3260      3270      3280      3290  
  
 2320      2330      2340      X  
 GGTGGGTG-----CCT--TAGCTCTTTCTCGATAGTTA----GAC  
 | | |    | |    | | | |    | | | |    | |  
 GGTGAGAGCAGACATCCTGACAGCTTTGAGACTACAGTGACAGTGACATTTACATGCAGACAGACAAACCA  
 3300      3310      3320      3330      3340      3350      3360  
  
 GGGTGGGGCCGTCCTGCAGAAGCGG  
 3370      3380      3390

5. ELLIS-012-FIG2AB.SEQ (1-2350)

XELAEIP      X.laavis amidating enzyme (AE-I) mRNA, complete cd

LOCUS      XELAEIP      2733 bp ss-mRNA      VRT      15-MAR-1989

DEFINITION      X.laavis amidating enzyme (AE-I) mRNA, complete cds.

ACCESSION      M18174

KEYWORDS amidating enzyme.  
 SOURCE X.laevis skin, cDNA to mRNA, clone pXAE457.  
 ORGANISM Xenopus laevis  
 Eukaryota; Animalia; Chordata; Vertebrata; Amphibia; Lissamphibia;  
 Anura; Archeobatrachia; Pipoidae; Pipidae; Xenopodinae.  
 REFERENCE 1 (bases 1 to 2733)  
 AUTHORS Mizuno,K., Ohsuye,K., Wada,Y., Fuchimura,K., Tanaka,S. and  
 Matsuo,H.  
 TITLE Cloning and sequence of cDNA encoding a peptide C-terminal  
 alpha-amidating enzyme from Xenopus laevis  
 JOURNAL Biochem. Biophys. Res. Commun. 148, 546-552 (1987)  
 STANDARD full automatic  
 COMMENT Amidating enzyme protein precursor is cleaved at two sites to  
 obtain the active enzyme.

FEATURES Location/Qualifiers  
 mRNA <1..2733  
 /note="AE-I mRNA"  
 sig\_peptide 266..376  
 /codon\_start=1  
 /note="amidating enzyme signal peptide"  
 mat\_peptide 377..1408  
 /codon\_start=1  
 /note="amidating enzyme"  
 CDS 266..1468  
 /note="amidating enzyme precursor"  
 /codon\_start=1  
 /translation="MASLSSSFLVLFLLFQNSCYCFRSPLSVFKRYEESTRSLNDCL  
 GTTRPVMSPGSSDYTLDIRMPGVPTPTESDTYLCKSYRLPVDDAYVVDFRPHANMDTA  
 HHMLLFGCNIPSSDDYHDCSAGTCMDKSSIMYAWAKNAPPTKLPEGVGRVCGKSGS  
 RYFVLQVHYGNVKAFOQKHKDCGTGTVRVTPKQPIAGIYLSMSVDVIPPGEEAVN  
 SDIACLYNRPTIHPFAYRVHHTQLGQVVSQFRVRHGKWSLIGRQSPQLPQAFYPVEHP  
 VEISPGDIIATRCFTGKGRTSATYIGGTSNDEMCLYIMYYMDAAHATSYMTCVQGTG  
 EPKLFQNIPEIANVPIPVSPDMMMMHGHGHHTEAEPEKNTGLQPKREEEEVLQGL  
 ITLGDSAV"

BASE COUNT 823 a 555 c 547 g 808 t  
 ORIGIN

Initial Score = 157 Optimized Score = 960 Significance = 8.29  
 Residue Identity = 46% Matches = 1185 Mismatches = 996  
 Gaps = 341 Conservative Substitutions = 0

```

                                     X      10      20
                                     ATGTCATGAAGTCTGAGTGG
                                     | ||| ||||| |
TCAGGAGTCCCTCTCTGTCTTTAAGAGGTATGAGGAATCTACCAGATCACTTTCCAATGACTGCTTGGAA
330      340      350      360      370      380      390      400

      30      40      50      60      70      80
ATAAACAGCACGGGATATCT-CTGTCTAAAGGAATATTACT----ACACCAGGA--AAAGGACACATTCGAC
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
CCACGCGGCGCGTTATGTCTCCAGGCTCATCAGATTATACTCTAGATATCCGCATGCCAGGAGTAATC--C
410      420      430      440      450      460      470

      90      100      110      120      130      140
AACAGGAAAGGA-GCCTGTACAGAAAACC--ACAGTGT-CCTGTGCATG-TGA----CATTTG-CCATGG
||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
GACAGAGTCGGACACATATTGTGCAAGTCTTACCGGCTGCCAGTGGATGATGAAGCCTATGTAGTTGACTT
480      490      500      510      520      530      540

150      160      170      180      190      200      210
GAAACAACCTGTTACAACGTGG-TGGT-CA-TTGTGCTGCTAGTGGGCTGTGAGAAGGTGGGAGCCGTGC
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
CAGACCAC-ATGCCAATATGGAATGCACATCAGATGCTTCTATTTGGATGCAATATACCTTCTTCCACTG
550      560      570      580      590      600      610

```

220 230 240 250 260 270 280







-CGGACGAA--GAGGACAGAGGAGACAGTCCGAAAAGTTATTTTCCGGCAAATCCTTTCCCTGTTTCGT  
 | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 GCTTTCAAATGGGGGTCA---CTGACCCCATCTAAAAAACAAATGCTCTGTAAGAC---TACAAATTTATT  
 2090 2100 2110 2120 2130 2140 2150

1780 1790 1800 1810 1820 1830  
 GACACTCCACCCCTTGTGGACACTTGAGTGTCA-TCCTTGCGC--CGGAAGGTCAGG--TGGTA--CCCGTC  
 | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 GTTACTGCTTTTAATTAGTAATGTTTCTATTAGGCCCTCCCTATTTCATATTCAAGCCTTTTATTCCAATC  
 2160 2170 2180 2190 2200 2210 2220

1840 1850 1860 1870 1880 1890 1900  
 --TGTAGGG--GC-GGGGAGACAGAGCCGCGGGGAGCTACGAGAATCGACTCACAGGGCGCCCGG-GCTT  
 | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 AGTGCATGTTGCTAGGGTAATTGGTACCC---TAGCAACCAG-ATC-CTAAAACTGCAAACTGGAGAAC  
 2230 2240 2250 2260 2270 2280

1910 1920 1930 1940 1950 1960  
 CGC-AAATGAAA--CTTTTT-----TAATCTCACAAGTTTCGTCGGGCTCGGCGACCTATGGCGTCGATC  
 | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 TGCTGAATAAAAGCTAAATAACAAAAAACACAA--ATAATAAAAAAT--GTAACCAACTGC--AAATT  
 2290 2300 2310 2320 2330 2340 2350

1970 1980 1990 2000 2010 2020 2030  
 CTTATTACCTTATCTGGCGCAAGATAAAACA-ACCAAAAGCCTTGACT--CCGGTACTAATTCTCCCTGC  
 | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 GTCAGAATATCACCCTG--TACAATCTACATCAGCTAAAAG--TTAATTTAAAGGT--GAACAACCCATA  
 2360 2370 2380 2390 2400 2410 2420

2040 2050 2060 2070 2080 2090  
 CGGCCCCCGTAAGCATAACGCGGCGATCTCCACTTTA-----AGAACCTGGCCGCTTCTGCCTGGTCT-C  
 | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 AGGAAGACATA--CAATTTGTGGATACACACTACAGACACTACAACCTAGATG-GCTCATTAAAGGAATAT  
 2430 2440 2450 2460 2470 2480

2100 2110 2120 2130 2140 2150 2160  
 GCTTTCGTAAACGGTCTTACAAAAGTAATTAGTT--CT--TGCTTTCAGCCTCCAAGCTTCTGCT-AGTCT  
 | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 GATTTACATTTTATTTATTAATAAATGAAATGATTTAACTGTTGATTT-TGAAT---TGATTATGTTGATTCT  
 2490 2500 2510 2520 2530 2540 2550

2170 2180 2190 2200 2210 2220 2230  
 ATGGCAGCATCAAGGCTGGTATTTGCTACGGCTGACCGCTACGCCCGCAATAAGGGTACTGGGCGGCCCCG  
 | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 AATGTTGAAT-----TGTTATTGGGTGCTGAAAAGTATCATAGGGTGGAAT--GTATACTTTTC-TCCTG  
 2560 2570 2580 2590 2600 2610 2620

2240 2250 2260 2270 2280 2290  
 ---TCSAAG-GCCCTTTGGT--TTCAGAAACCAAGGCCCCCTCATACCAACGTTTCG----ACT-TTGAT  
 | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 AGATTGGTGTGGTGTGGGTCTTACATAAATC--TTTACTTTGTACTATGATTTTTTCGAAAAACTCTTAAT  
 2630 2640 2650 2660 2670 2680 2690

2300 2310 2320 2330 2340 2350  
 TCTTGCCGGTACGTGGTGGTGGGTGCCTTAGCTCTTTCTCGATAG--TTAGAC  
 | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
 TAT----GTAACCTTCTTG--GAGTGAATAAAC-CITAAT--ATTGCATTGGG  
 2700 2710 2720 2730 X

6. ELLIS-012-FIG2AB.SEQ (1-2350)

HUMUT5094 Human chromosome 4 STS UT5094.

LOCUS HUMUT5094 468 bp ds-DNA PRI 28-MAY-1993  
 DEFINITION Human chromosome 4 STS UT5094.  
 ACCESSION U114477

KEYWORDS PCR primer; STS; microsatellite DNA; microsatellite marker;  
repeat polymorphism; sequence tagged site; tetranucleotide repeat.

SOURCE Homo sapiens DNA.

ORGANISM Homo sapiens  
Eukaryota; Animalia; Chordata; Vertebrata; Mammalia; Theria;  
Eutheria; Primates; Haplorhini; Catarrhini; Hominidae.

REFERENCE 1 (bases 1 to 468)

AUTHORS Gerken,S.C., Matsunami,N., Lawrence,E., Carlson,M., Moore,M.,  
Ballard,L., Melis,R., Robertson,M., Bradley,P., Elsner,T.,  
Tingey,A., Rodriguez,P., Albertsen,H., Lalouel,J.-M. and White,R.

TITLE Genetic and physical mapping of simple sequence repeat containing  
sequence tagged sites from the human genome

JOURNAL Unpublished (1993) See COMMENT for author address.

STANDARD full automatic

COMMENT Submitted by: Utah Center for Human Genome Research  
University of Utah, Dept. of Human Genetics  
2160 Eccles Institute of Human Genetics  
Salt Lake City, UT 84112  
e-mail: sts@corona.med.utah.edu

```
Initial Score      = 152  Optimized Score = 210  Significance = 7.95
Residue Identity  = 51%  Matches           = 262  Mismatches  = 173
Gaps              = 75   Conservative Substitutions = 0
```

```

660          670          680          690          700          710          720
TTCTGTGACTCCAGAGGGAGGACCAGGAGGGCACTCCTTGACAGGTCCTTACCTTGTTCTGGCGCTGACATC
          |||  |||  ||  ||
          ACTTGAGCCTGG----GA-GTC
          X      10

730          740          750          760          770          780          790          800
GG-CTTTGCTGCTGGCCCTGATCTTCATTACTCTCCTGTTCTCTGTGCTCAAATGGATCAGGAAAAAATTCC
||  |||  ||  |||  |||  ||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
GGAGGTTGCAG-TG-AGCTGA-CATCAT---GCCACTGCACTC-GAGC-CTGA-GCAACAG--AGCAAGACC
  20          30          40          50          60          70

```

010 020 070 080 050 040



/codon\_start=1  
/translation="MVSLEDFSRMYENVDPALASSPTRFHMMDFYSHNRQCLLQEKGI  
GTIHPYGTPLRTQHWSSSNHSIETQSTSSEEIVPSPSPPLPRIYKPCFVCQDKSSG  
YHYGVSACEGCKGFFRRSIQKNMVYTCHRDKNCIINKVTRNRQYCRLOKCFEVMGSK  
ESVRNDRNKKKKESPKPEAIESYILSPETQDLIEKVQKAHQETFPALCQLGKYTTSFS  
SEQRVSLDIDLWDFSELSTKCIKTVEFAKQLPGFTTLTIADQITLLKSACLDILIL  
RICTRYTPDQDTHTFSDGLTLNRTQMHNAGFGPLTDLVFAFANQLVPLEMDAETGLL  
SAICLICGDRQDLEQPDKVQDLQEPLEALKIYVTRRPQKPHMFPKMLMKITDLRSI  
SAKGAERVITLKEIPGAMPPLIQEMLENSEGLDTLGGGASSDAPVTPVAPGSCSPSL  
SPSSTHSSPSTHSP"

BASE COUNT      811 a      852 c      791 g      786 t  
ORIGIN

Initial Score    =    151    Optimized Score    =    961    Significance    =    7.88  
Residue Identity =    46%    Matches                =    1160    Mismatches    =    1053  
Gaps             =    295    Conservative Substitutions        =    0

```

                                     X      10
                                     ATGTCCATGAAGT---CTGAG
                                     ||  |  ||  |  |  |
TCTACAGTCACAACCGACAGTGCCTTTTGCAGGAGAAAGGATTGGGACCATTACCCGTACGGGACCCAC
  610      620      630      640      650  X  660      670

20      30      40      50      60      70      80
TGGATAACAGCAGCGGA-TATCTCTGTCTAAAGGAATATTACT-ACACCAGGA----AAAGGACACATTG
|  |  ||  |||  |||  |||  |  |  |  |  |||  |  |  |  |  |||  |  |||
TACGGACTCAACACTGGAGCAGCTCAACCACTCAATTGAGACTCAAAGCAGGAGTTCAGAGG--AGATTGT
  680      690      700      710      720      730      740

  90      100      110      120      130      140      150
ACAACAGGAAAGGA-GCCTGTACAGAAAACACAGTGTCTGTGCATGTGACATTTGCCATGGGAAACAA
||  |||  |  ||  |||  |  |||  |  ||  |  ||  |  ||  |  ||  |  ||  |
AC-CCAGCCCCCATCCCCACCACCGCTCCCCAGAACTCTACAAGCCCTGCTTTGTGT-GTCA-GGACAAGAG
  750      760      770      780      790      800      810

  160      170      180      190      200      210
CT---GTTACAACGTGGTGGTCA---TTGTG-CTGCTGCTAGTGGGCTGTGAGAAGGTGGGAG--CCGTGC
|  |  ||  ||  |  |||  ||||  |||  |  |||  |  ||||  |  ||  ||  |  |
TTGGGGTATCACTATGAGTCAGCGCTTGTGAAGGTTGCAA--GGGCTTT--TTCCGTGCGAGTATCCAGA
  820      830      840      850      860      870      880

220      230      240      250      260      270      280
AGAAC-TCCTGTGATAACTGTGAGCCTGGTACTTTCTGCAGAAAATAC-AATCCAGTCTGCAAGAGCTGCC
||||  |  |||  |  ||||  |  |  |||  |  |||  |  ||  |||  |||||
AGAACATGGTGT-ACACGTGTACAGAGACAAGAATTGCATCATAACAAAGTCACGC-GCAACCGCTGCC-
  890      900      910      920      930      940      950

290      300      310      320      330      340      350
TCCAAGTACCTTCTCCAGCATAGGTGGACAGCCGAAGTGAACATCTGCAGAGTGTGTGAGGCT-ATTTCA
|  |  |  |  |||  |  ||||  |  ||  |  |||  |  ||  |  ||  |  ||  |
-AGTATTGCCGATTGCAGAAATGTTTCGAGGTGGGAATG-----TCTAAAGAATCCGTACGGAATGATCGCA
  960      970      980      990      1000      1010

360      370      380      390      400      410      420
GGTTCAAGAAGTTTTGCTCTCTACCCACAACGCGGAGTGTGAG-TGCATTGAAGGATTCCATTG-CTTGGG
||||  |||  |  ||  |  |||  |  |||  |  |||  |  ||  |||  |  |
ACAAGAAGAAAAGGAGTCCCCAAAGCCTGAGGC-AATAGAGAGTTACAT--ACTGAGCCAGAGACACAAG
1020      1030      1040      1050      1060      1070      1080

430      440      450      460      470      480      490
GCCACAGTGCACAGATGTGAAAAGGACTGCAGG---CCTGGCCAGGAGCTAACGAAGCAGGGTTGCAAAAC
|  ||  ||  |  ||  ||||  |  |||  |||  |||  |  |||  |||  |||
ATCTCATTG-AGAAAGTGCAAAAGCCCACCAGGAGACCTCCCTGCA-CTCTGCCAGCTGG---GCAATA
1090      1100      1110      1120      1130      1140      1150

500      510      520      530      540      550      560
```

CTGTAGCTTGGGAACATTTA---ATGACCAGAACGGTACTGGCGTCTGTCCGAC--TG-GACGA----ACTG  
| | | | | | | | | | | | | | | | | | | | | |  
CACTA-----CAAGTTTTAGCTCGGAGCAGCGGGTTTCTCTGGAC-ATCGACCTGTGGGACAAGTTCAGTCT  
1160 1170 1180 1190 1200 1210

570 580 590 600 610 620 630  
CTCTCTAGACGGAAGGTCTGTGCTTAAGACCGGGACCACGGAGAAGGACGTGGTGTGTGGACCCCCTGTGGT  
| | | | | | | | | | | | | | | | | | | | | |  
AGCTCTCCACTAAA--TGTATCATCAAGACGGTGGAAATTGCCAA--AC-AGTTGCCGGGATTCACTACTCT  
1220 1230 1240 1250 1260 1270 1280

640 650 660 670 680 690  
GAGCTTC-TCTCCAGTACCACCATTTCTGTGA--CTCCAG-AGGGAGGACC--AGGAGGGCACTCCTTGCA  
| | | | | | | | | | | | | | | | | | | | | |  
GACCATCGCCGACCAG-ATCACCTCTCTGAAATCCGCCTGCCTGGATATTCTTATCCTGCGAAT--TTGCA  
1290 1300 1310 1320 1330 1340 1350

700 710 720 730 740 750 760  
GGTCTTA---CCTGTCTCTGGCGC--TGACATCGGCTTTGCTG--CTGGCCCTGATCTTCATTACTCTCC  
| | | | | | | | | | | | | | | | | | | | | |  
-CACGTTACACCCCTGATCAGGACCATGACAT--TCTCAGACGGACTGACCTAAACCGCACTCAGATGC  
1360 1370 1380 1390 1400 1410 1420

770 780 790 800 810 820  
TGTCTCTGTGCTCAAATGGATCA-GGA--AAAAATCCCCAC-ATATTCA---AGCAACCATTTAAGAAG  
| | | | | | | | | | | | | | | | | | | | | |  
ACAACGCGGGGTTTCGACCTCTCACAGACCTGGTCTTCGCTTCGCTAATCAGCTCGTGCCGCTTGAATGG  
1430 1440 1450 1460 1470 1480 1490

830 840 850 860 870 880 890  
ACCAC-TGGAG-CAGCTCAAGAGGAAGATGCTTGTAGCTGCCGATGTCCACAG-GAAGA--AGAAGGAGGAG  
| | | | | | | | | | | | | | | | | | | | | |  
ACGACGCTGAGACCGGTCTACTG--AG-TGC---CATCTGCC---TCATCTGTGGAGACCGGACGAGCTG  
1500 1510 1520 1530 1540 1550

900 910 920 930 940 950 960  
GAGGAGGC-TATGAGCTGTGATGTACTATCCTAGGAG-ATGTGTGGG--CCGAAACCGAGAAGCACTAGGAC  
| | | | | | | | | | | | | | | | | | | | | |  
GAGCAGCCAGATAAAGTG-GACAACT---GCAGGAGCCTCTTTTGAAGCGTTAAAGATCTACGTGAGGAC  
1560 1570 1580 1590 1600 1610 1620

970 980 990 1000 1010 1020 1030  
CCCACCATCTGTGGAACAGCACAAACCCACACCTGTTCTTACACATCATCCTAGATGATGTG--T  
| | | | | | | | | | | | | | | | | | | | | |  
CAGGCGA-CCCCAAAACCTCACATG-TTCCCCAAA---TGCTCATGAAGATCA-CAGACCTGCGGAGCAT  
1630 1640 1650 1660 1670 1680 1690

1040 1050 1060 1070 1080 1090 1100  
GGGCGCGCACCTCATCCAAGTCTCTTCTAACGCT-AACATATTTGCTTTTACCTTTTTTAATCTTTTTTTA  
| | | | | | | | | | | | | | | | | | | | | |  
CAGTGC-CAAGGGTGGGAGCGTGTGATCACTCTGAAGATGGAGATCCCGGGGCCATGCCCCCCTCATCC  
1700 1710 1720 1730 1740 1750 1760

1110 1120 1130 1140 1150 1160 1170  
AATTTAAATTTTATGTGTGTGAGTGTTTTGCTGCCTGTATGCACACGTGTGTGTGT---GTG-TGTGTGTG  
| | | | | | | | | | | | | | | | | | | | | |  
AGGAGATGTTGGAGAACTCGGAGGGGTTGGACACATTGGGGG----GTG-GGGCATCCAGTGATGCACCAG  
1770 1780 1790 1800 1810 1820

1180 1190 1200 1210 1220 1230  
ACACTCCTGATGCCTGAGGAGGTGAGAAGAAAGGGT-----TGGTTCCA-TAAGAACTGGAGTT--ATG  
| | | | | | | | | | | | | | | | | | | | | |  
TCACACAGTAGCACCAGGAAG-CTGCAGTCCCAGTCTGTCTCCAGTCCACTCACAGCAGCCCTCCACT  
1830 1840 1850 1860 1870 1880 1890







/sequenced\_mol="cDNA to mRNA"

/germline

BASE COUNT 905 a 279 c 371 g 751 t

ORIGIN

Initial Score = 150 Optimized Score = 808 Significance = 7.81  
Residue Identity = 46% Matches = 992 Mismatches = 847  
Gaps = 275 Conservative Substitutions = 0

320 330 340 350 360 370 380 390  
CCGAACTGTAACATCTGCAGAGTGTGTGCAGGCTATTTTCAGGTTCAAGAAGTTTTGCTCCTCTACCCACAAC  
|| || | || | || |  
GTATTTTCCATCAATTCATATC  
X 10 20

400 410 420 430 440 450 460  
GCGGAGTGTGAGTGCATTGAAGGATTCCATTGCTTGGGGCCACAGTGCACCAGAT-GTGAAAAGGACTGCAG  
| | | | | || | | | || || | | | | | || ||  
GTTTA-AATAATTTTTTCCA-AATGAAGTTGCTGTTGTTT-----TTTTTCATATCGTCTATCTTCTTCCTTCAG  
30 40 50 60 70 80

470 480 490 500 510 520  
GCCTGGCC-AGGAGCTAACGA-AGCAGGGTTGCAAAAC--CTGTA----GCTT-GGGAA--CATTTAATGAC  
|| | | | | | | | | | || | | | | || | | | |  
--CTGACCTCTGGGAAAAGGATATTAAATTTTGATAACATCATTAAACATCTTAAGGAAAGCAAATTATTGC  
90 100 110 120 130 140 150

530 540 550 560 570 580  
CAGAACGGTA---CTGGCGTCTGTGACCCCTGGACGAACT---GCT-CTCTAGACGGAAGGTCTGTGCTTAA  
| | | | | | | | | | | | | | | | | | | | |  
CTGAA-GATATCCCTCACGT-TTTAGAAAATGACATAATTATAGTTCCTCCTTATTTAA--TTTATAAAT-A  
160 170 180 190 200 210 220

590 600 610 620 630 640 650  
GACCGGGACCACGGA-GAAGGACGTGGT-GTGTGGACC-CCCTGTGGTGAGCTTCTCTCCAGTACCACCAT  
| | | | | | | | | | | | | | | | | | | | |  
TAAAGGAAAAATATATCACCTACATAATAATGTAGACCTTACTTTGATAAAC----CATCCTG-AAGAAGAT  
230 240 250 260 270 280 290

660 670 680 690 700 710 720  
TTCTGTGACTCCAGAGGGAGGACCAGGAGGGCA--CTCCTTGCAGGTCCT----TACCTTGTTCCTG-GCGC  
| | | | | | | | | | | | | | | | | | | | |  
TCATG-----CGATAAGGAAGA-AATTTGGGAATCCCATTTTCTCCTAAAGTACC-GGAACCGGAACGA  
300 310 320 330 340 350

730 740 750 760 770 780 790  
TGACATCGGCTTTGCTGCTGGCCCTGATCTTCATTACTCTCCTGTTCTCTGTGCTCAAATGGATCAGGAAAA  
|| | | | | | | | | | | | | | | | | | | | |  
CCACA-CGAACCAGAAATGGAACCTCAGGTTGAACCCGAACCAGGAC-CATTGCCTGAAGAGGTGAG--AGA  
360 370 380 390 400 410 420

800 810 820 830 840 850 860  
AATTCCTCCACATATTCAAGC--AACCA-TTTAAGAA-GACCACTGGAGCAGCTCAAGAGGAAGATGCTTGT  
| | | | | | | | | | | | | | | | | | | | |  
ACCTGAACCGGAACCAAGCAGAACCAAGAAAAAGAAATAGAAATGGA--AGAACAAGAAGAAG-TGATTGA  
430 440 450 460 470 480 490

870 880 890 900 910 920 930  
AGCTGCCGATGTCCACAGGAAGAAGGAGGAGGAGGCTATGAGCTGTGATGTAC----TATCCTAGG  
|| | | | | | | | | | | | | | | | | | | | |  
AGCTG-ATATG-GTATTAGACGAAGAACTGGAATTAATCCCTAAAAAGACA-GAACAAGATGTCCTAGA  
500 510 520 530 540 550 560

940 950 960 970 980 990  
AGATGTGTGGGCCGAAACCGAGAAGCACTAGGA---CCCCACCATCCTGTGGAA-CAGCAC---AAGCAACC  
| | | | | | | | | | | | | | | | | | | | |

A-----GTAAGCAAATTTTCGAGAAG-AATATGAATTACCTAACGT--TGTGGAATTAACCTCCTGAAG-AGAA  
570 580 590 600 610 620

1000 1010 1020 1030 1040 1050 1060  
CCACCACCCTGTTCTTACACATCATCCTAGATGATGTGTGGGC-GCGCACCTCATCCAAGTCTCTTCTAACG  
| | | | | | | | | | | | | | | | | | | | | |  
AGAAAAAATAATATTTTACATTTTGC-AGGTAATAAAAGTACAGCTTTCAATTTAAAAGAGATTATAAATT  
630 640 650 660 670 680 690

1070 1080 1090 1100 1110 1120 1130  
CTAACATATTTGTCTTTACCTTTTTTAAATCTTTTTTAAATTTAAATTTTATGTGTGTGAGTGTTCCT  
| | | | | | | | | | | | | | | | | | | | | |  
ATAAAAAAGATG---AAAGTTTAATGAATAGTTTATCTAGTT--CCTTTGATCATTTTTA-TACTCCTAAT  
700 710 720 730 740 750 760

1140 1150 1160 1170 1180 1190 1200  
GCCTGTATGCAC-ACGTGTGTGT-GTGTGTGT-GTGT-GACACTCCTGATGCCTGAGGAGGTCAGAAGAG-A  
| | | | | | | | | | | | | | | | | | | | | |  
G-TTG-AAGAACGAAATTTGAGTAATGCTTATAATGTGGATATTAATGAT-TATTATGATTTATTAAGAGCG  
770 780 790 800 810 820

1210 1220 1230 1240 1250 1260 1270  
AAGGGTTGGTTCCATAAGAACTGGAGTTATGGATGGCTGTGAGCCGNNNGATAGGTCGGGACGGAGACCTG  
| | | | | | | | | | | | | | | | | | | | | |  
TTACATATATTATGTAAGAC-GAAGATAATAA--GTTATATACAATAAATAGAATACCTAAAGATGAGTTA  
830 840 850 860 870 880 890

1280 1290 1300 1310 1320 1330  
TCTTCTTATTTTAA-CGTG--ACTGTA--TA-ATAAAA-----AAAAATGATATTCGGGAATTG--TAG  
| | | | | | | | | | | | | | | | | | | | | |  
TTATTTTTTCTTAAGCATGCATATGTAAATATATATAATTCTTTAAAAAATATATTTTATTAATGGCTAT  
900 910 920 930 940 950 960 970

1340 1350 1360 1370 1380 1390 1400  
AGATTGTCCTGACACCCTTCTAGTTAATGATCTAAGAGGAATTGTTGATACGTAGTATACTGTATATGTGTA  
| | | | | | | | | | | | | | | | | | | | | |  
A-ATTTGAAGA-----TTATATATATACCTCTGA---GAATT-TTACTTTAGATCAAATTTTTAAAGATTA  
980 990 1000 1010 1020 1030

1410 1420 1430 1440 1450 1460 1470  
TGTATATGTAT--ATGTAT-ATA-TAA-GACTCTTTTACTGTCAA-AGT-CAACCTAGAGTGTCTGGT-TAC  
| | | | | | | | | | | | | | | | | | | | | |  
TTTTTTTTTATCAAACGATGATACTAATGAAAAATGGTAGTTTAAATAATATAATCGAAAGTATAAAATATAT  
1040 1050 1060 1070 1080 1090 1100

1480 1490 1500 1510 1520 1530  
CAGGTCAATTTTATTG-GACATTTTACGTACACACACACACACACACACACACACGTT-----TATA  
| | | | | | | | | | | | | | | | | | | | | |  
CAAG--AAAGCTATAGAAAAATTAATGTAAAAAGAATAGA-AGAGAAGATAAAATATTTTTTTCAAATATA  
1110 1120 1130 1140 1150 1160 1170

1540 1550 1560 1570 1580 1590 1600  
CTACGTAC-TGTTATCGGTATTCTACGTCATATAATGGGATAGGGTAAAAGGAAACCAAAGAGTGAGTGATA  
| | | | | | | | | | | | | | | | | | | | | |  
TGAGGTACATGCTTTTG--ATTTTAAAT--TATTACATTATATATTTTCTCGAAATC----AGTTA-TTAAA  
1180 1190 1200 1210 1220 1230

1610 1620 1630 1640 1650 1660 1670  
TTATTGTGGAGGTGACAGACTACCCCTTCTGGGTACGTAGGGACAGACCTCCTTCGGACTGTCTAAACTCC  
| | | | | | | | | | | | | | | | | | | | | |  
TTATAGTGAAGATGATTTAATTCACC-ACGCAGTAGATA-TTATGAATATTACAAGGA-TAGATATATCTCC  
1240 1250 1260 1270 1280 1290 1300

1680 1690 1700 1710 1720 1730 1740  
CCTTAGAAGTCTCETCAAGTTCCCGGACGAAGAGGACAGAGGAGACACAGTCCGAAAAGTTATTT-TTCCGG  
| | | | | | | | | | | | | | | | | | | | | |

```

-----AAGGGTAAT--ATCTGCCCTTATTTA-TGTATTTCTTAAATA-AG---GTAAATATATTTCTTGAC
      1310      1320      1330      1340      1350      1360

1750      1760      1770      1780      1790      1800      1810
CA-AATCCTTTCCCTG-TTTTCG---TGACACTCCACCCCTTGTGGA-CAC-TTGAGTGTATCCTTGCGCCG
|| ||| | | | | ||| | || | ||| ||| ||| | | | | |
CACAATATGTAACAAGAAATAGAAATGACTTTACATTATTATTGGAGCACAAATGA-TTTATTATTAAGGACA
      1370      1380      1390      1400      1410      1420      1430

      1820      1830      1840      1850      1860      1870
GA-----AGGTC-AGGTGCT--ACCCGTCTGTAGGGCGGGGAGACAGCCGCGGGGAGCTACGAGAA
|| | | | | | | | | | | | | | | | | | | | | | |
GAGATAATATATCGAGATTTTTTAAACATTTTTTAAAGC-ATAAAACA---CCTC-ATGTACATTTGAAAA
      1440      1450      1460      1470      1480      1490      1500

1880      1890      1900      1910      1920      1930      1940
TCGACTCACAGGGCGCCCGGGCTTCGCAATGAACTTTTTAATCTCACAAGTTTCG---TCCG-GGCTC
|| || | | | | | | | | | | | | | | | | | |
AAAACAATCATG---ATAATGCATACCAATT--AGTTCCGTGGTCT--CAAGTATTGTTTTCTGAATTTA
      1510      1520      1530      1540      1550      1560

      1950      1960      1970      1980      1990      2000
GGCGGACCTATGGC--GTCG-ATCCTTATTACCTTATC--CTGGCGCCAAGAT--AAAACAACC-----AA
| ||||| | | | | | | | | | | | | | | | | | | |
ATAGTACCTATGACTTTTCGAATTTTAAATATATGATCATATTGTTTCATGATTCTTATCATGCTTTTGTAG
      1570      1580      1590      1600      1610      1620      1630

      2010      2020      2030      2040      2050      2060      2070
A-AGCCTTGACTCCGGTACTAATTCTCCCTGCCGGCCCCGTAAGCATAACGCGGCGATCTCCACTTTAAGA
| |||| | | | | | | | | | | | | | | | | | |
ATTATTTTGAAGGAGATGATAAAT-TAAATCAAGTATTGAATGATAATAAGAGAATAAGACCA--TTGATA
      1640      1650      1660      1670      1680      1690      1700

      2080      2090      2100      2110      2120      2130
ACCTGGCCGCGTTCTGCCTGGTCTCGCTTTTCTGTAAG-CGGTTCT---TACAAAA-GTA--ATTAGTTCTTGC
| | | | | | | | | | | | | | | | | | | | | |
ATTTAGATAAATTTT--TTAAT---GAGTTATTAATTTGTTTTTAATAGAAAAATGTAGGAATA-TTATCGA
      1710      1720      1730      1740      1750      1760      1770

2140      2150      2160      2170      2180      2190      2200      2210
TTTCAGCCTCCAAGCTTCTGCTAGT-CTATGGCAGCATCAAGGCTGGTATTTGCTACGGCTGACCGCTACGC
| | | | | | | | | | | | | | | | | | | | |
CATTGG--AAGAATATTTTGAAGTACTGTAAACAAGTTGGACGAGTACT---TTCAGATGATCATGACAT
      1780      1790      1800      1810      1820      1830

      2220      2230      2240      2250      2260      2270      2280
CGCCGCAATAAGGGTACTGGGCGGCCCGTCAAGGCCCTTTGGTTTCAGAAACCAAG-GCCCCCTCATAC
| | | | | | | | | | | | | | | | | | | | |
TGATTCTGA-AAGAATA-----TACGGAGAA--TTATTTTACTCCAGAAGAAGAGCAAGCTTTAAAA
      1840      1850      1860      1870      1880      1890      1900

      2290      2300      2310      2320      2330      2340
CAACGTTTCGACTTTGATTCTTGCCGGTACGTGGTGGTGGTGCCTTA-GCTCTT--TCTCGATAGTTA---
| | | | | | | | | | | | | | | | | | | | |
GATTTTAAAGAAGATGAGGCTT--CTGTACATCAT-TTGGAT-CATTACGATAATAATATGGATCCTTATTA
      1910      1920      1930      1940      1950      1960

2350
-GAC
||
TGAATATCAAGGCGAATTTGCTAGTTATGAAGAAGAAGAGGATGAACTCAGAAA
      1970 X      1980      1990      2000      2010      2020

```

LOCUS ACLRGAL 1508 bp ds-DNA BCT 15-MAR-1990  
 DEFINITION A.laidlawii 16S ribosomal RNA small subunit gene.  
 ACCESSION M23932  
 KEYWORDS 16S ribosomal RNA; ribosomal RNA small subunit.  
 SOURCE A.laidlawii (strain JA1) DNA.  
 ORGANISM Acholeplasma laidlawii  
 Prokaryotae; Tenericutes; Mollicutes; Mycoplasmas; Mycoplasmatales;  
 Acholeplasmataceae.  
 REFERENCE 1 (bases 1 to 1508)  
 AUTHORS Weisburg,W.G., Tully,J.G., Rose,D.L., Petzel,J.P., Oyaizu,H.,  
 Yang,D., Mandelco,L., Sechrest,J., Lawrence,T.G., van Etten,J.L.,  
 Maniloff,J. and Woese,C.R.  
 TITLE A phylogenetic analysis of the mycoplasmas: Basis for their  
 classification  
 JOURNAL J. Bacteriol. 171, 6455-6467 (1989)  
 STANDARD full automatic  
 COMMENT Draft entry and computer-readable sequence [1] kindly submitted by  
 C.R.Woese, 19-JAN-1989.  
 FEATURES Location/Qualifiers  
 rRNA 1..1480  
 /note="16S ribosomal RNA small subunit (3' end approx.)"  
 BASE COUNT 433 a 303 c 433 g 339 t  
 ORIGIN

Initial Score = 146 Optimized Score = 568 Significance = 7.53  
 Residue Identity = 46% Matches = 679 Mismatches = 633  
 Gaps = 161 Conservative Substitutions = 0

```

910      920      930      940      950      960      970
TGAGCTGTGATGTACTATCCTAGGAGATGTGTGGGCCGAAACCGAGAAGCACTAGGACCCACCATCCT-GT
                                     ||      |  |  |  |
                                     TTTATATGGAGAGTTTGATCCT
                                      X      10      20

980      990      1000      1010      1020      1030      1040
GGAACAGCACAAGCAACCCACACCTGTTCTTACACAT-CATCCTAGATGATGTGTGGCGCGCACCTCA
||  |||  |  |||  |  |||  |  |||  ||  |||  ||  |||  ||  |||  ||
GGCTCAG---GATGAACGCTGGCGGCGTG-CCTAATACATGCAAGTCGAACGAAGCATCTTCGGATGCTTAG
      30      40      50      60      70      80      90

1050      1060      1070      1080      1090      1100      1110
T--CCAAGTCTCTTCTAACGCTAACAT-ATTGTCTTTACCTTTTTTAAATCTTTTTTAAATTTAAATTTT
|  |  |  |  |||  |  |||  |  |||  ||  |||  ||  |||  ||  |||  ||
TGGCGAACGGGTGAGTAACACGTAGATAACCTACCTTTAACTCGAGGATAACTCCGGGAACTGGAGCTAAT
      100      110      120      130      140      150      160

1120      1130      1140      1150      1160      1170      1180
ATGTGTGTGAGTGTGTTTGCCTGCCTG-TATGCACACGTGTGTGTGTGTGTGTGTGACA---CTC-CTGAT
|  |  |  |  |||  |  |||  |  |||  ||  |||  ||  |||  ||  |||  ||
A-CTG-GATAG-GATGTG--TGCATGAAAAAACACATTTAAAGATTTATCGGTTTAAAGAGGGGTCTGCGGC
      170      180      190      200      210      220

1190      1200      1210      1220      1230      1240      1250
GCCTGAGGAGGTGAGAAGAGAAAGGTTGGTTCCATAAGAACTGGAGTTATGGATGG-CTGTGA-GCCGGNN
||  |||  ||  |  |  |||  |  ||  |||  ||  ||  ||  ||  |||  ||  ||
GCAITAGTTAGTTGGTGGGTAAGAG--CTACC--AAGACGATGAATCGTAGCCGGAAGTCTGAGAGGTCTACC
230      240      250      260      270      280      290

1260      1270      1280      1290      1300      1310      1320
NGATAGGTCGGGACGGAGACCTGTCTTCTTATTTTAACTGACTGTATAATAAAAAAATGATATTTTCGG
|  |  |  |||  |||  |||  |  |  |||  ||  ||  ||  ||  |||  |||  |||
GGCCACATTGGGACTGAGAACGGCC--CAAACCTCTACGGGA-GGCA---GCAGTAGGGAAT----TTTCGG
300      310      320      330      340      350

1330      1340      1350      1360      1370      1380      1390

```

GAATTGTAGAGATTGCTCGAC--ACC--CTTCTAGTTAATGATCTAAGAGGAATTGTTGATACGTAGTATA  
||| | ||| ||||| | | | | ||| | | | | | | | | | | | | | | |  
CAATGGGGGAAA---CCCTGACCGAGCAACGCCGCTGAACGA-CGA--AGTACT--TCGGTATGTAAAGTT  
360 370 380 390 400 410 420

1400 1410 1420 1430 1440 1450 1460  
C-TGTATATGTGTATGTATATGTATATATAAGACTCTTTTACT-GTCAAAGTCAACCTAGAGTGTG  
| | ||||| | | | | ||| | | | | | | | | | | | | | | |  
CTTTTATATG-GGAAGAAAAATTAA--AATTGACGGTACCATATGAATAAGCCCCGGCTAACTA-TGTGGC  
430 440 450 460 470 480 490

1470 1480 1490 1500 1510 1520 1530  
TGGTTACCAGGTCAATTTTATTGGACATTTTACGTCACACACACACAC---ACACACA-CACACACACGT--  
| | | ||| | || ||| ||| | | | ||| | | | | | | | | | |  
AGCAGCCGCGGTAATACATAGGGGGC---GAGCGTTATCCGGATTACTGGGCGTAAAGGGTGCCTAGGTGG  
500 510 520 530 540 550 560

1540 1550 1560 1570 1580 1590 1600  
TTATACTACGTACTGTTATCGGTATTCTACGTCATATAATG-GGATAGGGTAAAAGGAAACCA-AAGAGTGA  
|||| | | ||| | || | | | | ||| | | | |||||  
TTATAAAAGTTTGTGGTGAAGTGCAGTGCTTAACGCTGTGAGGCTATG--AAACTATATACTAGAGTGA  
570 580 590 600 610 620 630

1610 1620 1630 1640 1650 1660  
GTGATATTATTGTGGAGGTGACAGACTACCCCT--TCTGGGT--ACGTAGGGACAGACCTCCTTCGGACTGT  
| | | |||| | | ||| | ||| | |||| | | | | | | | |  
GACAGAGGCAAGTGAATTCCATGTGTAGCGGTAAATGCGTAAATATATGGA-GGAACACC-----AGTGG  
640 650 660 670 680 690

1670 1680 1690 1700 1710 1720 1730  
CTAAAC-TCCCCTTAGAAGTCTCGTCAAGTTCCCGGACGAAGAGGACAGAGGAG-ACACAGTCCGAAAAGT  
| | | | | | | | | |||| | | |||| | |||| | |||| | |  
CGAAGCGCGCTTGCTGGGTCTATACTGACACTGATGCACGAA-AG-CGTGGGGAGCAAACAG---GA----T  
700 710 720 730 740 750

1740 1750 1760 1770 1780 1790 1800 1810  
TATTTTTCGGCAAATCCTTTCCCTGTTTCGTGACACTCCACCCCTTGTGGACACTTGAGTGTATCCTTGC  
|| | ||| | ||| ||| ||| ||| ||| ||| ||| ||| |||  
TAGATACCCTGGTAGTCCACGCCGTAAACGATGAGA---ACTAAGTGTGGC-CATAAG-GTCAGTGTGC  
760 770 780 790 800 810 820

1820 1830 1840 1850 1860 1870  
GCCGGAAG-GTCAGGTGGTACCCGTCTGTAG--GGCGGGGAGACAGAGCCGGGGGAGCT-ACGAGAATC  
| | | ||| | ||| | | | | | | | | | | | | | |  
AGTTAACGCATTAAGTTCTCCGCCTGAGTAGTACGTACGCAAGTATGAACTCAAAGGAATTGACGGGACCC  
830 840 850 860 870 880 890

1880 1890 1900 1910 1920 1930 1940  
GACTCACAGGGCGCCCGGGCTTCGCAAAT-GAA---ACTTTTTTAATCTCA-CAAGTTTCGTC-----CGG  
| | | ||| | ||| ||| ||| ||| ||| ||| ||| ||| |||  
CGCACAAAGCGGTGGATCATGTTGTTAATTGGAAGATACACGAAAAACCTTACCAGGTCTTGACATACTCTG  
900 910 920 930 940 950 960

1950 1960 1970 1980 1990 2000 2010  
GCTCGGCGGACCTATGGCGTC-GATCCTTA-TTACCTTATCCTGGCGCCAAGATAAAACAACCAAAGCCTT  
||| | || ||| ||| | | ||| ||| | | | | | | | | | |  
CAAAGGCTTAGAAATAAGTTCCGAGGCTAACAGATGTACAGGTGGTGCACGGTTGTCTCAGCTCGTGTCTG  
970 980 990 1000 1010 1020 1030 1040

2020 2030 2040 2050 2060 2070  
GA---CTCCGGTACTAATTCTCCCTGC---CGGCCCGGTAAGCATAACGCGGCGATC-TCCACTTTAAGAA  
|| | ||| ||| ||| | | ||| ||| | | | | | | | | | |  
GAGATGTTGGGT--TAAGTCCCGCAACGAGCGCAACCTTATTGCTA--GTTACCATCATTAAGTTGGGGAC  
1050 1060 1070 1080 1090 1100

2080 2090 2100 2110 2120 2130 2140



DRLHTIRHYQHVLAVDPEKAAQMSQVHTLHVIEERRNQSLSLLYKVPYVAQEIQEE  
IDELLQEQRADMDQFTASISETPVDVRVSSESEEIIPFHPFHPALPENEGSCVGE  
QDGGGLIGAEKVINSKNKVDENMVIDETLDVKEMIFNAERVGGLEERESVGPLREDF  
SLSSSALIGLLVIAVAIATVIVISLVMLRKRQYGTISHGIVEVDPMLTPEERHLNKMQ  
NHGYENPTYKYLEX"

BASE COUNT 941 a 804 c 913 g 865 t  
ORIGIN

Initial Score = 145 Optimized Score = 772 Significance = 7.47  
Residue Identity = 49% Matches = 959 Mismatches = 721  
Gaps = 273 Conservative Substitutions = 0

```

      490      500      510      520      530  X  540      550
GCAGGGTTGCAAAACCTGTAGCTTGGGAACATTTAATGACCAGAACGGTACTGGCGTCTGTGACCCCTGGAC
                                     |  |  |  |  |  |  |
                                     CCAAACTACGT--GCGGTGTGC
                                     X      10      20

      560      570      580      590      600      610      620
GAACTGCTCTCTAGACGGAAGGTCTGTGCTTAAGA-CCG-GGACCACGGAGA-AGGACGTGGTGTGTGGACC
||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
TAA--GC-GAGGAGTCCGAGTGTGTGAGCTTGAGAGCCGCGCTAGAGCGACCCGCGAGGGATGGCGGCC
      30      40      50      60      70      80

      630      640      650      660      670      680      690
CCCTGTGGTGAGCTTCTCTCCAGTACCACCATTTCTGTGACTCCAGAGGGAGGACCAGGAGGGCACTCCTT
||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
ACC--GGGACGCGGCGCGCCGAG--CCAC--GGGCAG--GCTCCTGCTTCTGCTGCTGGTGGG-GCTCAGG
90      100      110      120      130      140      150

      700      710      720      730      740      750      760
GCAGGTCCTTACCTTGTCTCTGGC--GCTGACATC--GGCT-TTGCTGC---TGGCCCTGATCTTCATTACT
||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
GC--GCC--TGCTTGGCGCTGGCCGGCT-ACATCGAGGCTCTTGAGCCAATG--CCGGAACAGGATT---
      160      170      180      190      200      210

      770      780      790      800      810      820
CTCCTGTTCTCTGTG-CTCAAATGG--AT---CAGGAAAAAATTCCCCACATATTCAAGCAACCATTTAAG
||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
-TGCTGTT-GCTGAGCCTCAAATCGCAATGTTTGTGGGAAGTT-----AAATATGCATGTGAACATTGAGA
      220      230      240      250      260      270

      830      840      850      860      870      880      890
AAGACCACTGG--AGCAGCT-CAAGAGGAAGATGCTTGTAGCTGCCGATGTCCACAGGAAGAAGAAGGAGGA
|  |  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
CTGGGAAATGGGAACCTGATCCAACAGGCA---CCAAG-AGCTG-CTTTG-AAACA-AAAGAAGAA-GTTCT
280      290      300      310      320      330      340

      900      910      920      930      940      950      960
GGAGGA-GGCTATGAGCTGTGAT---GTACTA-TCCTAGGAGATGTG-TGG-GCCGAAACCGAGAAGCACTA
||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
TCAGTACTGTGAGGAGATGT-ATCCAGAGCTACAGATCACAAATGTGATGGAGGCAAAACGCGGGTTAGTA
      350      360      370      380      390      400      410

      970      980      990      1000      1010      1020      1030
-GGACCCCAACATCCTGTGGAACAGCACAAGCAACCCACCACTGTTCTTACACATCATCCTAGATGATG
||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
TTGACAACCTGGTGCCGGAGGGACA---AAAAGCAATGCAAGAGTGGCTTTGTACAC--CTTTCAAG-TGTCT
      420      430      440      450      460      470      480

      1040      1050      1060      1070      1080      1090      1100
TGTGGGCGGCACCTCATCCAAGTCTCTTCTAACGCTAACATATTTGTCTTTACCTTTTTTAAATCTTTTTT
|||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
CGTGGGTGAATTTGTAAGTGATGTC-CTGCTAGTTC-CAGAAAAGTGCC---AGTTTTCCACA-----
      490      500      510      520      530      540
```



1110 1120 1130 1140 1150 1160  
TAAATTTAAATTTTATGTGTGTGAGTGTTCCTGATGC--ACACGTGTGTGTGTGTGTGT  
| | | | | | | | | | | | | | | | | | | | | | | | | | | |  
AAGAGCGGATGGAGGTGTGTGAGAATCACCAGCACTGGCACACGGTAGTCAAAGAGGCATGTCTGACTCAGG  
550 560 570 580 590 600 610

1170 1180 1190 1200 1210 1220 1230  
GTGTGACACTCCTGATGCCTGAGG-AGG---TCAGAAGAGAAAGGGTTG---GTTCCATAAGAACTGGAG  
| | | | | | | | | | | | | | | | | | | | | | | | | | | |  
GAATGAC-CTTAT-ATAGCTACGGCATGCTGCTCCCATGTG---GGGTAGACCAGTTCAT-GGCACTGAA-  
620 630 640 650 660 670

1240 1250 1260 1270 1280 1290  
TTATGGATG--GCTGTGAG-CCGGNNGATAGG-TCGG-----GACGGAGACCTGTCTTCTTATTTTAA  
| | | | | | | | | | | | | | | | | | | | | | | | | | | |  
-TATGTGTGCTGCCCTCAGACAAAGATTATTGGATCTGTGTCAAAGAAGAGGAAGAGGAAGATGAAGAGGA  
680 690 700 710 720 730 740

1300 1310 1320 1330 1340 1350  
CGTG--ACTGTATAATAA---AAAAAATGATATTTCCGGGAATTGTAGAGATTGTCCTGACACCCTTCTA  
| | | | | | | | | | | | | | | | | | | | | | | | | | | |  
AGAGGAAGAGGAAGATGAAGAGGAAGACTATGATGTTT--ATAAAAGTGAATTTCTACTGAAGCAGATCTG  
750 760 770 780 790 800 810

1360 1370 1380 1390 1400 1410 1420  
GTTAATGA--TCTAAG-AGGAATTGTTGATACGTAGTATATACTGTATATGTGTATGTATATGTATATGT-ATA  
| | | | | | | | | | | | | | | | | | | | | | | | | | | |  
G--AA-GACTTCACAGAAGCAGCTGTGGAT---GAGGATGATGAGGATGAGGAAGAAGGGGAGGAAGTGGTG  
820 830 840 850 860 870 880

1430 1440 1450 1460 1470 1480 1490  
TATAA--GACTCTTTTAC--TGTCAAAGTCAACCTAGAGTGTCTGGTTACCAGGTCAATTTTATTGGACATT  
| | | | | | | | | | | | | | | | | | | | | | | | | | | |  
GAGGACCGAGATTACTACTATGACACCTTCAAAGGAGA-TGACT-ACAATGAGGAGAATCCTACTGAACCCG  
890 900 910 920 930 940 950

1500 1510 1520 1530 1540 1550 1560  
TTA-CGTC-ACACACACACACACACACACA-CACACACGTTTATACTACGTACTGTTATCGGTATTCT-ACG  
| | | | | | | | | | | | | | | | | | | | | | | | | | | |  
GCAGCGACGGCACCATGTGACACAAGGAAATTACTCATGAT--GTCAAAGTTCCTCCAAC--TCCTCTGCCA  
960 970 980 990 1000 1010 1020

1570 1580 1590 1600 1610 1620  
TCATATAATG-GGATAGGGTAAAAGGAAACCAAGAGTGAGTGAT-ATTATTGTG-GAGGTGACAGA---CT  
| | | | | | | | | | | | | | | | | | | | | | | | | | | |  
ACCAATGATGTTGAT-GTGATTTTCGAGACCTCTG-CAGA-TGATAATGAGCATGCTCGCTCCAGAAGGCT  
1030 1040 1050 1060 1070 1080 1090

1630 1640 1650 1660 1670 1680 1690  
A--CCCCTTCTGGGTACGTAGGGACAG--ACCTCCTTCGGACTGTCTAAAACTCCCCTTAGAAGTCTCGTCA  
| | | | | | | | | | | | | | | | | | | | | | | | | | | |  
AAGGAGCAGCTGGAGA-TTCGGCACCGCAACCGAAT--GGACAGGGTAAAGAAGGAATGGGAAG---AGGCA  
1100 1110 1120 1130 1140 1150

1700 1710 1720 1730 1740 1750 1760  
-AGTTCCCGGACGAAGA----GGACAGAGGAGACACAGTCCGAAAAGTTATTTTCCGGCAAATCCTTTCCC  
| | | | | | | | | | | | | | | | | | | | | | | | | | | |  
GAGCTTCAAG-CTAAGAACCTCCCAAAGCAG--AGAGGCAG-ACTCTGATT----CAGCACTTCCAAGCCA  
1160 1170 1180 1190 1200 1210 1220

1770 1780 1790 1800 1810 1820  
TGTTTCGTGACACTCCA-CCCCTTGTGGACACTTGAGTGTATC--CTTGCG-----CCGGAAGGTGAG-GT  
| | | | | | | | | | | | | | | | | | | | | | | | | | | |  
TGGTTAAAG-CTTTACAGAAGGAAGCAGCCA-GTGAGAAGCAGCAGCTGGTGGAGACCCACCTGGCCCGAGT  
1230 1240 1250 1260 1270 1280 1290



OS Homo sapiens (human)  
 OC Eukaryota; Animalia; Metazoa; Chordata; Vertebrata; Mammalia;  
 OC Theria; Eutheria; Primates; Haplorhini; Catarrhini; Hominidae.  
 XX  
 RN [1]  
 RP 1-3523  
 RA Vostrov A.A., Quitschke W.W., Schwarzman A.L., Blangy A.,  
 RA Cuzin F., Wesley U.V., Hagag N.G., Goldgaber D.;  
 RT "Cloning of a protein that binds to a recognition sequence in the  
 RT APP promoter";  
 RL Unpublished.

XX  
 FH Key Location/Qualifiers  
 FH  
 FT CDS <69..>2154  
 FT /product="binding protein"  
 FT /note="putative"  
 FT /codon\_start=1  
 FT source 1..3523  
 FT /organism="Homo sapiens"  
 FT /dev\_stage="adult"  
 FT /sequenced\_mol="cDNA to mRNA"  
 FT /tissue\_type="brain"  
 XX

SQ Sequence 3523 BP; 941 A; 804 C; 913 G; 865 T; 0 other;

Initial Score = 145 Optimized Score = 772 Significance = 7.47  
 Residue Identity = 49% Matches = 959 Mismatches = 721  
 Gaps = 273 Conservative Substitutions = 0

```

      490      500      510      520      530  X   540      550
GCAGGGTTGCAAAACCTGTAGCTTGGGAACATTTAATGACCAGAACGGTACTGGCGTCTGTGACCTGGAC
                                     |   |   |   |   |
                                     CCAAACACTACGT--GCGGTGTGC
                                     X      10      20

      560      570      580      590      600      610      620
GAAGTCTCTCTAGACGGAAGGTCTGTGCTTAAGA-CCG-GGACCACGGAGA-AGGACGTGGTGTGTGGACC
||  ||      ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
TAA--GC-GAGGAGTCCGAGTGTGTGAGCTTGAGAGCCGCGCGCTAGAGCGACCCGGCGAGGGATGGCGGCC
      30      40      50      60      70      80

      630      640      650      660      670      680      690
CCCTGTGGTGAGCTTCTCTCCAGTACCACCATTTCTGTGACTCCAGAGGGAGGACCAGGAGGGCACTCCTT
||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
ACC--GGGACCGCGCGCGCCGAG--CCAC--GGGCAG--GCTCCTGCTTCTGCTGCTGGTGGG-GCTCAGG
90      100      110      120      130      140      150

      700      710      720      730      740      750      760
GCAGGTCCTTACCTTGTTCCTGGC--GCTGACATC--GGCT-TTGCTGC---TGGCCCTGATCTTCATTACT
||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
GC--GCC--TGCCTTGGCGCTGGCGCGCT-ACATCGAGGCTCTTGCAGCCAATG--CCGGAACAGGATT---
      160      170      180      190      200      210

      770      780      790      800      810      820
CTCCTGTTCTCTGTG-CTCAATGG--AT---CAGGAAAAAATTCCCCACATATTCAAGCAACCATTTAAG
|  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
-TGCTGTT-GCTGAGCCTCAATCGCAATGTTTTGTGGGAAGTT-----AAATATGCATGTGAACATTGAG
      220      230      240      250      260      270

      830      840      850      860      870      880      890
AAGACCACTGG--AGCAGCT-CAAGAGGAAGATGCTTGTAGCTGCCGATGTCCACAGGAAGAAGAAGGAGGA
|  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
CTGGGAAATGGGAACCTGATCCAACAGGCA---CCAAG-AGCTG-CTTTG-AAACA-AAAGAAGAA-GTTCT
280      290      300      310      320      330      340
  
```

900 910 920 930 940 950 960  
GGAGGA-GGCTATGAGCTGTGAT---GTACTA-TCCTAGGAGATGTG-TGG-GCCGAAACCGAGAAGCACTA  
| | | | | | | | | | | | | | | | | | | |  
TCAGTACTGTGCAGGAGATGT-ATCCAGAGCTACAGATCACAATGTGATGGAGGCCAACCGCGGGTTAGTA  
350 360 370 380 390 400 410

970 980 990 1000 1010 1020 1030  
-GGACCCCACCATCCTGTGGAACAGCACAAAGCAACCCACCACCTGTTCTTACACATCATCCTAGATGATG  
| | | | | | | | | | | | | | | | | | | |  
TTGACAACTGGTGCCGGAGGGACA--AAAAGCAATGCAAGAGTCGCTTTGTACAC--CTTTCAAG-TGTCT  
420 430 440 450 460 470 480

1040 1050 1060 1070 1080 1090 1100  
TGTGGCGCGCACCTCATCCAAGTCTCTTCTAACGCTAACATATTGTCTTTACCTTTTTTAAATCTTTTTT  
| | | | | | | | | | | | | | | | | | | |  
CGTGGGTGAATTTGTAAGTGATGTC-CTGCTAGTTC-CAGAAAAGTGCC---AGTTTTCCACA-----  
490 500 510 520 530 540

1110 1120 1130 1140 1150 1160  
TAAATTTAAATTTTATGTGTGTGAGTGTTTTGC-CTG----CCTGTATGC--ACACGTGTGTGTGTGTGT  
| | | | | | | | | | | | | | | | | | | |  
AAGAGCGGATGGAGGTGTGTGAGAATCACGAGCACTGGCACACGGTAGTCAAAGAGGCATGTCTGACTCAGG  
550 560 570 580 590 600 610

1170 1180 1190 1200 1210 1220 1230  
GTGTGACACTCCTGATGCCTGAGG-AGG----TCAGAAGAGAAAGGGTTG----GTTCCATAAGAACTGGAG  
| | | | | | | | | | | | | | | | | | | |  
GAATGAC-CTTAT-ATAGCTACGGCATGCTGCTCCCATGTG---GGGTAGACCAGTTCAT-GGCACTGAA-  
620 630 640 650 660 670

1240 1250 1260 1270 1280 1290  
TTATGGATG--GCTGTGAG-CCGGNNNGATAGG-TCGG-----GACGGAGACCTGTCTTCTATTTTAA  
| | | | | | | | | | | | | | | | | | | |  
-TATGTGTGCTGCCCTCAGACAAAGATTATTGGATCTGTGTCAAAAGAAGAGGAAGAGGAAGATGAAGAGGA  
680 690 700 710 720 730 740

1300 1310 1320 1330 1340 1350  
CGTG--ACTGTATAATAAA-----AAAAAATGATATTTCGGGAATTGTAGAGATTGTCCTGACACCCCTCTA  
| | | | | | | | | | | | | | | | | | | |  
AGAGGAAGAGGAAGATGAAGAGGAAGACTATGATGTTT--ATAAAAGTGAATTTCTACTGAAGCAGATCTG  
750 760 770 780 790 800 810

1360 1370 1380 1390 1400 1410 1420  
GTTAATGA--TCTAAG-AGGAATTGTTGATACGTAGTATACTGTATATGTGTATGTATATGTATATGT-ATA  
| | | | | | | | | | | | | | | | | | | |  
G--AA-GACTTCACAGAAGCAGCTGTGGAT---GAGGATGATGAGGATGAGGAAGAAGGGGAGGAAGTGGTG  
820 830 840 850 860 870 880

1430 1440 1450 1460 1470 1480 1490  
TATAA--GACTCTTTTAC--TGTCAAAGTCAACCTAGAGTGTCTGTTTACCAGGTCAATTTTATTGGACATT  
| | | | | | | | | | | | | | | | | | | |  
GAGGACCGAGATTACTACTATGACACCTTCAAAGGAGA-TGACT-ACAATGAGGAGAATCCTACTGAACCCG  
890 900 910 920 930 940 950

1500 1510 1520 1530 1540 1550 1560  
TTA-CGTC-ACACACACACACACACACA-CACACACGTTTATACTACGTACTGTTATCGGTATTCT-ACG  
| | | | | | | | | | | | | | | | | | | |  
GCAGCGACGGCACCATTGTCAGACAAGGAAATTACTCATGAT--GTCAAAGTTCCTCCAAC--TCCTCTGCCA  
960 970 980 990 1000 1010 1020

1570 1580 1590 1600 1610 1620  
TCATATAATG-GGATAGGGTAAAGGAAACCAAAGAGTGAGTGAT-ATTATTGTG-GAGGTGACAGA---CT  
| | | | | | | | | | | | | | | | | | | |  
ACCAATGATGTTGAT-GTGTATTTGAGACCTCTG-CAGA-TGATAATGAGCATGCTCGCTTCAGAGGCT  
1030 1040 1050 1060 1070 1080 1090

1630 1640 1650 1660 1670 1680 1690  
A--CCCCCTTCTGGGTACGTAGGGACAG--ACCTCCTTCGGACTGTCTAAAACTCCCCTTAGAAGTCTCGTCA  
| | | | | | | | | | | | | | | | | | | | | |  
AAGGAGCAGCTGGAGA-TTCGGCACCGCAACCGAAT--GGACAGGGTAAAGAAGGAATGGGAAG---AGGCA  
1100 1110 1120 1130 1140 1150  
  
1700 1710 1720 1730 1740 1750 1760  
-AGTTCCTCCGGACGAAGA----GGACAGAGGAGACACAGTCCGAAAAGTTATTTTCCGGCAAACTCCTTTCCG  
| | | | | | | | | | | | | | | | | | | | | |  
GAGCTTCAAG-CTAAGAACCTCCCCAAAGCAG--AGAGGCAG-ACTCTGATT----CAGCACTTCCAAGCCA  
1160 1170 1180 1190 1200 1210 1220  
  
1770 1780 1790 1800 1810 1820  
TGTTTCGTGACACTCCA-CCCCTTGTGGACACTTGAGTGTATC--CTTGCG-----CCGGAAGGTGAG-GT  
| | | | | | | | | | | | | | | | | | | | | |  
TGGTTAAAG-CTTTAGAGAAGGAAGCAGCCA-GTGAGAAGCAGCAGCTGGTGGAGACCCACCTGGCCCGAGT  
1230 1240 1250 1260 1270 1280 1290  
  
1830 1840 1850 1860 1870 1880 1890  
GGTACCCTG-CTGTAGGGGGCGGGGAGACAGA-GCCGCGGGGAGCTACGAGAATCGACTCACAGGGCGCCCC  
| | | | | | | | | | | | | | | | | | | | | |  
GGAAGCTATGCTGAATGACC--GCCGTCGGATGGCTCTGGAGAACTACCTGGCT-GCCTTGCACTCTGACCC  
1300 1310 1320 1330 1340 1350 1360  
  
1900 1910 1920 1930 1940 1950 1960  
GGGCTTCG--CAAATGAAACTTTTTTAATCTCAC-AAGTTTCGTCCGGGCTCGGCGGACCTATGGCGTCGAT  
| | | | | | | | | | | | | | | | | | | | | |  
--GCCACGGCCTCATCGCATTCTCCAGGCCTTACGGCGTTATGTCCGTGCT--GAGAACAAA-----GAT  
1370 1380 1390 1400 1410 1420  
  
1970 1980 1990 2000 2010 2020 2030  
C-CTTA-TTACCTTATCCTGGCGCCAAGATAAAACAACCAAAAGC-CTTGACTCCGGTACTAATTC-TCCCT  
| | | | | | | | | | | | | | | | | | | | | |  
CGCTTACATACC--ATC----CGTCATTA-CCAGCATGTGTTGGCTGTTGAC-CCAGAA--AAGCGGCCCCA  
1430 1440 1450 1460 1470 1480  
  
2040 2050 2060 2070 2080 2090  
GCCGCCCCCGTAAGCATAACGCGCGCATCTCCAC----TTAAGAACCTGGCCGCGTTCTGCCTGGTCTC-  
| | | | | | | | | | | | | | | | | | | | | |  
GATGAAATCCCAGGTGATGACAC----ATCTCCACGTGATTGAAGAAAGGAGGAAC-CAAAGCCT-CTCTCT  
1490 1500 1510 1520 1530 1540  
  
2100 2110 2120 2130 2140 2150 2160  
GCTTTCGTAA--ACGGT-TCTTACAAAAGTAATT--AG---TTCCTGCTTTCAGCCTCCA--AGCTTC-TGC  
| | | | | | | | | | | | | | | | | | | | | |  
GCTCTACAAAGTACCTTATGTAGCCCAAGAAATTCAAGAGGAAATTGATGAGCTCCTTCAGGAGCAGCGTGC  
1550 1560 1570 1580 1590 1600 1610 1620  
  
2170 2180 2190 2200 2210 2220  
TAGTCTATGG--CAGCATCAAGGCTGGTATTT--GCTACGGCTGACCGC-TACGCCGCGCAATAAGGGTAC  
| | | | | | | | | | | | | | | | | | | | | |  
-AG-ATATGGACCAG-TTCACTGCCTCAATCTCAGAGACCCCTGTGGACGTCCGGGTGAGCTCTGAGGAGAG  
1630 1640 1650 1660 1670 1680 1690  
  
2230 2240 2250 2260 2270 2280 2290  
TG-GGCGGCCCGTGAAGGCCCTTTGGTTTCA-GAAACCCAAGGCCCCCTCA-TACCAACGTT-TCGACTT  
| | | | | | | | | | | | | | | | | | | | | |  
TGAGGAGATCC--CACCGTTCCACCCCTTCCACCCCTTCCAGCCCTACCTGAGAACGAAGGATCTGGAGTG  
1700 1710 1720 1730 1740 1750 1760  
  
2300 2310 2320 2330 2340 X  
TGATTCTTGCCGGTACGTGG--TGGTGGGTGCC--TTAGCTCTTTCGATAGTTAGAC  
| | | | | | | | | | | | | | | | | | | | | |  
GGA---GAGCAGGATGGGGGACTGATCGGTGCCGAAGAGAAAGTGATTAACAGTAAGAATAAAGTGGATGAA  
1770 1780 1790 1800 1810 1820

AACATGGTCATTGACGAGACTCTGGATGTTAAGGAAA  
1830 1840 1850 1860

12. ELLIS-012-FIG2AB.SEQ (1-2350)

HSB15RNA Homo sapiens mRNA for HB15

LOCUS HSB15RNA 1761 bp RNA PRI 31-JUL-1992  
DEFINITION Homo sapiens mRNA for HB15  
ACCESSION Z11697  
KEYWORDS HB15 gene; immunoglobulin superfamily.  
SOURCE human  
ORGANISM Homo sapiens  
Eukaryota; Animalia; Metazoa; Chordata; Vertebrata; Mammalia;  
Theria; Eutheria; Primates; Haplorhini; Catarrhini; Hominidae.  
REFERENCE 1 (bases 1 to 1761)  
AUTHORS Zhou,L., Schwarting,R., Smith,H.M. and Tedder,T.F.  
TITLE A novel cell-surface molecule expressed by human interdigitating  
reticulum cells, Langerhans cells and activated lymphocytes that is  
a new member of the immunoglobulin superfamily  
JOURNAL J. Immunol. 149, 735-742 (1992)  
STANDARD full automatic  
REFERENCE 2 (bases 1 to 1761)  
AUTHORS Tedder,T.F.  
TITLE Direct Submission  
JOURNAL Submitted (11-FEB-1992) T.F. Tedder, Division of Tumor Immunology,  
Dana-Farber Cancer Institute/Harvard Medical School, 44 Binney St.,  
Boston, MA, 02115-6084, USA  
STANDARD full automatic  
COMMENT \*source: tissue=human tonsil;  
\*source: cell\_type=lymphocyte;  
\*source: clone\_library=cDNA library in lambda gt-11; \*source:  
clone=pHB15;  
\*source: is\_macronuclear=N;  
\*source: is\_proviral=N;  
\*source: is\_germline=N.  
FEATURES Location/Qualifiers  
sig\_peptide 11..67  
mat\_peptide 68..625  
/product="HB15"  
/note="proposed amino-terminus of mature protein product"  
polyA\_signal 1248..1253  
CDS 11..628  
/evidence=EXPERIMENTAL  
/note="a cell-surface molecule expressed by  
interdigitating reticulum cells, Langerhans cells and  
activated lymphocytes.A member of the immunoglobulin  
superfamily"  
/product="HB15"  
/codon\_start=1  
/translation="MSRGLQLLLLSCAYSLAPATPEVKVACSEDVDLPCTAPWDPQVP  
YTVSWVKLLEGGEERMETPQEDHLRGQHYHQKGQNGSFDAPNERPYSLKIRNTTSCNS  
GTYRCTLQDPDQGRNLSGKVILRVGTGCPAQKKEETFKKYRAEIVLLALVIFYLTLLI  
FTCKFARLQSFDPDFSKAGMERAFLPVTSPNKHGLVTPHKTELV"  
BASE COUNT 453 a 399 c 442 g 467 t  
ORIGIN

Initial Score = 143 Optimized Score = 697 Significance = 7.33  
Residue Identity = 47% Matches = 843 Mismatches = 719  
Gaps = 207 Conservative Substitutions = 0

X 10 20  
ATGTCC-ATGAACTGCTGAGTG

||| || ||

TGAGCTGCGCCTACAGCCTGGCTCCCGCAGCGCGGAGGTGAAGGTGGCTTCCGAAGATGTGGACTTGC

40 50 60 70 80 90 100 110

30 40 50 60 70 80  
GATAACAGCAC---GGGATATCTCTGT---CTA-AAGGAATATTACTACACCAGGAAAAGGACACATTTCGA  
| | | | | | | | | | | | | | | | | | | | | |  
CCTGCACCGCCCCCTGGGATCCGCAGGTTCCCTACACGGTCTCCTGGGTCA--AGTTATTGGAGGGTGGTGA  
120 130 140 150 160 170 180

90 100 110 120 130 140 150  
CAACAGGAAAGGAGCCTGTACAGAAAACCA--CAGTGTCTGTGCATGTGACATTTGCGCATGGGAAACA  
| | | | | | | | | | | | | | | | | | | | | |  
AGAGAGG-ATGGAGACACCCAGGAAGACCACCTCAGGGGAC--AGCA-CTATCAT--CAGAAGGGGCAA-A  
190 200 210 220 230 240

160 170 180 190 200 210 220  
ACTGTTACAACGTGGTGGTCATTGTGCTGCTGTAGTGGGCTG-TGA---GAA--GGTGGGAGCCGTGCAGA  
| | | | | | | | | | | | | | | | | | | | | |  
ATGTTCTTTTCGACGCCCCCAATGAAAGGC-CCTA-TTCCCTGAAGATCCGAAACACTACCAGC--TGC--A  
250 260 270 280 290 300 310

230 240 250 260 270 280  
ACTC-CTGTGATACTGT-CAGCCTG--GTAC---TTTCTGCAGAAAATACAATCCAGTCTGCAAGAGCTGC  
| | | | | | | | | | | | | | | | | | | | | |  
ACTCGGGGACATACAGGTGCACTCTGCAGGACCCGGATGGGCAGAGAAACCTA---AGT-GGCA--AGGTGA  
320 330 340 350 360 370

290 300 310 320 330 340 350  
CCTCCA-AGT-AC--CTTCTCCAGCATAG-GTGGACAGCCGAACTGTAACATCTGCAGAGTGTGTGCAGG--  
| | | | | | | | | | | | | | | | | | | | | |  
TCTTGAGAGTGACAGGATGCCTGCACAGCGTAAAGAGAGACTTTTAAGAAATACAGAGCG-GAGATTGTC  
380 390 400 410 420 430 440

360 370 380 390 400 410 420  
CTATTTCAGGTTCAAGAAGTTTTGTCTCTACCCACA--ACGCGGAGT-GTGAGTGCATTGAAGGATTCCA  
| | | | | | | | | | | | | | | | | | | | | |  
CTGCTGCTGGCTCTGGTTATTTT-CTACTTAACACTCATCATTTTCACTTGTAAAGT---TTGCACGGCTACA  
450 460 470 480 490 500 510

430 440 450 460 470 480  
TTGCTTGGGGCCACAGTGCACCAGATGTGAAAAGGACTGCAGGCCT--GGCCAG-GAGCT----AACGAAGC  
| | | | | | | | | | | | | | | | | | | | | |  
GAGTATCTTCCAGATTTTCTAAAGCTGGCATGGAACG-AGCTTTTCTCCAGTTACCTCCCCAAATAAGC  
520 530 540 550 560 570 580

490 500 510 520 530 540 550  
----AGGG-TTGCAAAACCTGTAGCTTGGGAACATTTAATGACCAGAACGGTACTGGCGTCTGTGACCCTG  
| | | | | | | | | | | | | | | | | | | | | |  
ATTTAGGGCTAGTGACTCTCACAAGACAGAAC-TGCTATGAGCAGGA--TTTCTGCAGGTTCTTCTTCTCTG  
590 600 610 620 630 640 650

560 570 580 590 600 610 620  
GACGAACCTGCTCTCTAGACGGAAGGTCTGTGCTTAAGACCGGGACCACGGAGAAGGA-CGT--GGTG-TGTG  
| | | | | | | | | | | | | | | | | | | | | |  
AAGCTGAGGCTC---AG-GGGTGTGCTGTCTGTTACACTGGAGGAGAGAAGAATGAGCCTACGCTGAAGAT  
660 670 680 690 700 710 720

630 640 650 660 670 680 690  
GACCCCTGTGGTGAGCTTCTCTCCAGTACCACCATTT--CTGTGACTCCAGAGGGAGGACCAGGAGGGCA  
| | | | | | | | | | | | | | | | | | | | | |  
GGCATCCTGTGAAGTCCTTAC-CTCACTGAAAACATCTGGAAGGGGATCCCACCCCATTTTC-TGTGGGCA  
730 740 750 760 770 780 790

700 710 720 730 740 750  
CTCCTTGCAGGTCCTTACCT-TGTTCTGGCGCTGA--CATCGGCT-TTGCTGC-TGGC--CCTGATCTTCA  
| | | | | | | | | | | | | | | | | | | | | |  
GGCCTCGAAAACCATCACATGACCACATAGC-ATGAGGCCACTGCTGCTTCTCCATGGCCACCTTTTCAGCG  
800 810 820 830 840 850 860

760 770 780 790 800 810 820  
TTACTCTCCTGTTCTCTGTGCTCAAATGGATCAGGAAAAAATCCCCACATATTCAAGCAACCATTTAAGA  
| | | | | | | | | | | | | | | | | | | | | | | | | | | |  
AT-GTATGCAG--CTATCTGGTCAA---CCTCCTGGACATTTTTTTCAGTCATATAAAAGCTA--TGGTGAGA  
870 880 890 900 910 920

830 840 850 860 870 880 890  
AGACCACTGGAGCA-GCTCAAGAGGAAGATGCTTGTAGCTGCCGATGTCCACAGGAAGAAGAAGGAGGAGGA  
| | | | | | | | | | | | | | | | | | | | | | | | | | | |  
TG-CAGCTGGAAAAGGGTCTTGGGAAATATGAATG--CCCCAGCTGGCCGTGACAGACTCTGAGGA-CA  
930 940 950 960 970 980 990

900 910 920 930 940 950  
GGAGGC----TATG-AGCTGTGATGTACTATCCTAGGA---GATGTG-TGGGCCG---AAACCGAGAAGCAC  
| | | | | | | | | | | | | | | | | | | | | | | | | | | |  
GCTGTCTCTTCTGCATCT-TGGGGACATCTCTTTGAATTTTCTGTGTTTTGCTGTACCAGCCAGATGTTT  
1000 1010 1020 1030 1040 1050 1060

960 970 980 990 1000 1010 1020  
TAGGACCCACCATCTGTGGAACAGCACAAAGCAACCCACCAACCTG---TTCTTA-CACATCATCCTAGA  
| | | | | | | | | | | | | | | | | | | | | | | | | | | |  
TACGTCTGGGAGAAATTG----ACAGATCAAGCTGTGAGA-CAGTGGGAAATATTTAGCAAATAAT--TTCC  
1070 1080 1090 1100 1110 1120 1130

1030 1040 1050 1060 1070 1080 1090  
TGATGTG-TGGGCGCGC-ACCTCATCCAAGT-CTCT-TCTAAGCTAACATATTTGTC-TTTACCTTTTTTA  
| | | | | | | | | | | | | | | | | | | | | | | | | | | |  
TGGTGTGAAGGTCTGCTATTACTAAGGAGTAATCTGTGTACAAAGAAATAACAAGTCGATGAACTATTCCC  
1140 1150 1160 1170 1180 1190 1200

1100 1110 1120 1130 1140 1150  
AATC----TTTTTTTAAATTTAAATTTTATGTGT-GTGA-GTGTTTTGCCTGCCTGTATGCACACGTGTGTG  
| | | | | | | | | | | | | | | | | | | | | | | | | | | |  
CAGCAGGGTCTTTTCATCTGGGAAAGACATCCATAAAGAAGCAATAAAGAAGAGTG---CCACATTTATTTT  
1210 1220 1230 1240 1250 1260 1270

1160 1170 1180 1190 1200 1210 1220  
TGTGTGTGTGTGTGACACTC-CTGA-TGCCTGAGGAGGTGAG-AAGAGAAAGGGTTGGTTCCATA-AGAACT  
| | | | | | | | | | | | | | | | | | | | | | | | | | | |  
TATATCTATGTACTTTGTCAAAGAAGGTTTGTGTTTTTCTGCTTTTGAAA--TCTGTATCTGTAGTGAGAT  
1280 1290 1300 1310 1320 1330 1340

1230 1240 1250 1260 1270 1280 1290  
GGAGTTATGGA-TGGCTGTGAGCCGNNNGATAGGTCGGGACGGAGACCTGTCTTCTTATTTTAACGTGACT  
| | | | | | | | | | | | | | | | | | | | | | | | | | | |  
AGCATTGTGAACTGACAGGCAGCCTG---GACA--TAGAGAGGGAGA--AGAAGTCAGA---GAGGGTGACA  
1350 1360 1370 1380 1390 1400

1300 1310 1320 1330 1340 1350 1360  
GTATA-ATAAAAAAAAAATGATATTCGGGAAT--TGTAGAGATTGTCCTGACACCCTTCTAGTTAA--TGA  
| | | | | | | | | | | | | | | | | | | | | | | | | | | |  
AGATAGAGAGCTATTTAATGGCCGCTGGAATGCTGGGCTGACGGTGACGCTGGGTGCTCGTCCACTTGT  
1410 1420 1430 1440 1450 1460 1470

1370 1380 1390 1400 1410 1420 1430  
TCTA--AGAGGAATTGTTGATACGTAGTATACTGTATAT-GTGATG-TATATGTATATGTATA-TATAAGA  
| | | | | | | | | | | | | | | | | | | | | | | | | | | |  
CCCACTATCTGGGTGCATGATCTTGAGCAAGTTCCTTCTGGTGTCTGCTTTCTCCAT-TGTAAACCACAAGG  
1480 1490 1500 1510 1520 1530 1540

1440 1450 1460 1470 1480 1490 1500  
CTCTTTTACTGTCAAAGTCAACCTACAGTGTCTGGTTACCAGGTCAATTTTATTGG---ACATTTTACGTCA  
| | | | | | | | | | | | | | | | | | | | | | | | | | | |  
CTGTTGCATGGGCTAA-TGA---AGA---TC---ATA-TACGTGAAAATCTTTGAAAACATATAAAG-CA  
1550 1560 1570 1580 1590 1600



```

      1510      1520      1530      1540      1550      1560
C-ACACACACAC-ACACACACACACACGTTTATACTACGTACTGTTATCGGTATTCTACG--TCATATAA
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
CTATACAGATTGAACTC-CATTGAGTC-ATTATCCTTGCTA-TGATGATGGTGTGTTTGGGGATGAGAGGG
1610      1620      1630      1640      1650      1660      1670

1570      1580      1590      1600      1610      1620      1630
TGGGAT--AGGGTAAAGGAAACCAAGAGTGAGTGATATTATTGTGGAGGTGACAGACTACCCCTTCTGGG
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
TGCTATCCATTCTCATGTTTTCC----ATTGTTTGAAACAA---AGAAGGTTACCAAGAAGCCTTTCCTGT
1680      1690      1700      1710      1720      1730      1740

1640      1650      X      1670      1680      1690      1700
TACGTAGGGACAGACCTCCTTCGGACTGTCTAAACTCCCCTTAGAAGTCTCGTCAAGTCCCGGACGA
| | | | | | | | | |
AGCCTTCTGTAGGAATTC
      1750      1760

```

### 13. ELLIS-012-FIG2AB.SEQ (1-2350)

S53354 B-cell activation protein=B-G antigen IgV domain h

```

LOCUS      S53354      2574 bp      PRI      05-APR-1993
DEFINITION B-cell activation protein=B-G antigen IgV domain homolog [human,
            SAC-activated B lymphocytes, Genomic/mRNA, 2574 nt]
ACCESSION  S53354
KEYWORDS   .
SOURCE      human SAC-activated B lymphocytes
ORGANISM    Unclassified.
            Unclassified.
REFERENCE   1 (bases 1 to 2574)
AUTHORS     Kozlow,E.J., Wilson,G.L., Fox,C.H. and Kehrl,J.H.
TITLE       Subtractive cDNA cloning of a novel member of the Ig gene
            superfamily expressed at high levels in activated B lymphocytes.
JOURNAL      Blood 81, 454-461 (1993)
STANDARD    full automatic
COMMENT      This entry [NCBI gibbsq 123744] was created by the journal scanning
            component of NCBI/GenBank at the National Library of Medicine.
            This sequence comes from Fig. 7 and 1.
FEATURES             Location/Qualifiers
     mRNA             271..2574
     CDS               312..929
                     /note="B-G antigen IgV domain homolog; For the protein
                     sequence (NCBI gibbsq 123745): Method: conceptual
                     translation supplied by author. This sequence comes from
                     Fig. 1."
                     /product="B-cell activation protein"
                     /codon_start=1
                     /translation="MSRGLQLLLLSCAYSLAPATPEVKVACSEDVDLPCTAPWDPQVP
                     YTVSWVKLLEGGEERMETPQEDHLRGQHYHQGQNGSFDAPNERPYSLKIRNTTSCNS
                     GTYRCTLDQPDGQRNLSGKVILRVTCGPAQRKEETFKKYRAEIVLLALVIFYLTLLI
                     FTCKFARLQSFDFSKAGMERAFLPVTSPNKHGLVTPHKTELV"
BASE COUNT      649 a      595 c      689 g      641 t
ORIGIN

```

```

Initial Score      =      143  Optimized Score      =      897  Significance      =      7.33
Residue Identity   =      46%  Matches               =      1088  Mismatches       =      951
Gaps               =      279  Conservative Substitutions      =      0

```

```

                                     X      10      20
                                     ATGTCC-ATGAAGTCTGAGTG
                                     ||| | || ||
TGAGCTGCGCCTACAGCCTGGCTCCCGGACGCCGGAGGTGAAGGTGGCTTGCTCCGAAGATGTGGACTTGC
340      350      360      370      380      390      400      410

```

30 40 50 60 70 80  
GATAAACAGCAC---GGGATATCTCTGT---CTA-AAGGAATATTACTACACCAGGAAAAGGACACATTTCGA  
| | | | | | | | | | | | | | | | | | | | | |  
CCTGCACCGCCCCCTGGGATCCGCAGGTTCCCTACACGGTCTCCTGGGTCA--AGTTATTGGAGGGTGGTGA  
420 430 440 450 460 470 480

90 100 110 120 130 140 150  
CAACAGGAAAGGAGCCTGTCTACAGAAAACCA---CAGTGTCTGTGCATGTGACATTTGCGCATGGGAAACA  
| | | | | | | | | | | | | | | | | | | | | |  
AGAGAGG-ATGGAGACACCCAGGAAGACCACCTCAGGGGAC--AGCA-CTATCAT--CAGAAGGGGCAA-A  
490 500 510 520 530 540

160 170 180 190 200 210 220  
ACTGTTACAACGTGGTGGTCAATTGTGCTGCTAGTGGGCTG-TGA---GAA--GGTGGGAGCCGTGCAGA  
| | | | | | | | | | | | | | | | | | | | | |  
ATGGTTCTTTTCGACGCCCCCAATGAAAGGC-CCTA-TTCCCTGAAGATCCGAAACACTACCAGC--TGC--A  
550 560 570 580 590 600 610

230 240 250 260 270 280  
ACTC-CTGTGATACTGT-CAGCCTG--GTAC---TTTCTGCAGAAAATACAATCCAGTCTGCAAGAGCTGC  
| | | | | | | | | | | | | | | | | | | | | |  
ACTCGGGGACATACAGGTGCACTCTGCAGGACCCGGATGGGCAGAGAAACCTA---AGT-GGCA--AGGTGA  
620 630 640 650 660 670

290 300 310 320 330 340 350  
CCTCCA-AGT-AC--CTTCTCCAGCATAG-GTGGACAGCCGAAGTGAACATCTGCAGAGTGTGTGCAGG--  
| | | | | | | | | | | | | | | | | | | | | |  
TCTTGAGAGTGACAGGATGCCCTGCACAGCGTAAAGAAGAGACTTTTAAGAAATACAGAGCG-GAGATTGTC  
680 690 700 710 720 730 740

360 370 380 390 400 410 420  
CTATTTTCAGGTTCAAGAAGTTTTGCTCCTCTACCCACA--ACGCGGAGT-GTGAGTGCATTGAAGGATTCCA  
| | | | | | | | | | | | | | | | | | | | | |  
CTGCTGCTGGCTCTGTTATTTT-CTACTTAACACTCATCATTTTCACTTGTAAAGT---TTGCACGGCTACA  
750 760 770 780 790 800 810

430 440 450 460 470 480  
TTGCTTGGGGCCACAGTGCACCAGATGTGAAAAGGACTGCAGGCCT--GGCCAG-GAGCT----AACGAAGC  
| | | | | | | | | | | | | | | | | | | | | |  
GAGTATCTTCCAGATTTTCTAAAGCTGGCATGGAACG-AGCTTTTCTCCAGTTACCTCCCAATAAGC  
820 830 840 850 860 870 880

490 500 510 520 530 540 550  
----AGGG-TTGCAAAACCTGTAGCTTGGGAACATTTAATGACCAGAACGGTACTGGCGTCTGTCGACCCTG  
| | | | | | | | | | | | | | | | | | | | | |  
ATTTAGGGCTAGTGACTCCTCACAAGACAGAAC-TGGTATGAGCAGGA--TTTCTGCAGGTTCTTCTTCTCTG  
890 900 910 920 930 940 950

560 570 580 590 600 610 620  
GACGAACTGCTCTCTAGACGGAAGGTCTGTGCTTAAGACCGGGACCACGGAGAAGGA-CGT--GGTG-TGTG  
| | | | | | | | | | | | | | | | | | | | | |  
AAGCTGAGGCTC---AG-GGGTGTGCCTGTCTGTACACTGGAGGAGAGAAGAATGAGCCTACGCTGAAGAT  
960 970 980 990 1000 1010 1020

630 640 650 660 670 680 690  
GACCCCTGTGGTGAGCTTCTCTCCAGTACCACCATTT--CTGTGACTCCAGAGGGAGGACCAGGAGGGCA  
| | | | | | | | | | | | | | | | | | | | | |  
GGCATCCTGTGAAGTCTTTCAC-CTCACTGAAAACATCTGGAAGGGGATCCACCCCATTTTC-TGTGGGCA  
1030 1040 1050 1060 1070 1080 1090

700 710 720 730 740 750  
CTCCTTGCAGGTCTTACCT-TGTTCTGGCGCTGA--CATCGGCT-TTGCTGC-TGGC--CCTGATCTTCA  
| | | | | | | | | | | | | | | | | | | | | |  
GGCCTCGAAAACCATCACATGACCACATAGC-ATGAGGCCACTGCTGCTTCTCCATGGCCACCTTTTCAGCG  
1100 1110 1120 1130 1140 1150 1160

760 770 780 790 800 810 820  
TTACTCTCCTGTTCTCTGTGCTCAAATGGATCAGGAAAAATTCCCCACATATTTCAAGCAACCATTTAAGA  
| | | | | | | | | | | | | | | | | | | | | |  
AT-GTATGCAG--CTATCTGTGCAA--CCTCCTGGACATTTTTTCAGTCATATAAAAGCTA--TGGTGAGA  
1170 1180 1190 1200 1210 1220 1230

830 840 850 860 870 880 890  
AGACCACTGGAGCA-GCTCAAGAGGAAGATGCTTGTAGCTGCCGATGTCCACAGGAAGAAGAAGGAGGAGGA  
| | | | | | | | | | | | | | | | | | | | | |  
TG-CAGCTGGAAAAGGGTCTTGGGAAATATGAATG--CCCCAGCTGGCCCGTACAGACTCCTGAGGA-CA  
1240 1250 1260 1270 1280 1290

900 910 920 930 940 950  
GGAGGC----TATG-AGCTGTGATGTACTATCCTAGGA---GATGTG-TGGGCCG---AAACCGAGAAGCAC  
| | | | | | | | | | | | | | | | | | | | | |  
GCTGTCCTCTTCTGCATCT-TGGGGACATCTCTTTGAATTTTCTGTGTTTTGCTGTACCAGCCAGATGTTT  
1300 1310 1320 1330 1340 1350 1360

960 970 980 990 1000 1010 1020  
TAGGACCCACCATCTGTGGAACAGCACAAAGCAACCCACCCTG---TTCTTA-CACATCATCCTAGA  
| | | | | | | | | | | | | | | | | | | | | |  
TACGTCTGGGAGAAATTG----ACAGATCAAGCTGTGAGA-CAGTGGGAAATATTTAGCAAATAAT--TTCC  
1370 1380 1390 1400 1410 1420 1430

1030 1040 1050 1060 1070 1080 1090  
TGATGTG-TGGGCGCGC-ACCTCATCCAAGT-CTCT-TCTAACGCTAACATATTTGTC-TTTACCTTTTTTA  
| | | | | | | | | | | | | | | | | | | | | |  
TGGTGTGAAGGTCCTGCTATTACTAAGGAGTAATCTGTGTACAAAGAAATAACAAGTCGATGAACCTATTTCC  
1440 1450 1460 1470 1480 1490 1500

1100 1110 1120 1130 1140 1150  
AATC----TTTTTTTAAATTTAAATTTATGTGT-GTGA-GTGTTTTGCCTGCCTGTATGCACACGTGTGTG  
| | | | | | | | | | | | | | | | | | | | | |  
CAGCAGGGTCTTTTCATCTGGGAAAGACATCCATAAAGAAGCAATAAAGAAGAGTG---CCACATTTATTTT  
1510 1520 1530 1540 1550 1560 1570

1160 1170 1180 1190 1200 1210 1220  
TGTGTGTGTGTGTGACTC-CTGA-TGCCTGAGGAGGTCAG-AAGAGAAAGGGTTGTTCCATA-AGAACT  
| | | | | | | | | | | | | | | | | | | | | |  
TATATCTATATGTACTTGTCAAAGAAGGTTTGTGTTTTTCTGCTTTTGAAA--TCTGTATCTGTAGTGAGAT  
1580 1590 1600 1610 1620 1630 1640

1230 1240 1250 1260 1270 1280 1290  
GGAGTTATGGA-TGGCTGTGAGCCGNNNGATAGGTCGGGACGGAGACCTGCTTCTTATTTTAACGTGACT  
| | | | | | | | | | | | | | | | | | | | | |  
AGCATTGTGAAGTGCAGGCAGCCTG---GACA--TAGAGAGGGAGA--AGAAGTCAGA---GAGGGTGACA  
1650 1660 1670 1680 1690 1700

1300 1310 1320 1330 1340 1350 1360  
GTATA-ATAAAAAAAAAATGATATTTCGGGAAT--TGTAGAGATTGCTGACACCCTTCTAGTTAA--TGA  
| | | | | | | | | | | | | | | | | | | | | |  
AGATAGAGAGCTATTTAATGGCCGGCTGGAATGCTGGGCTGACGGTGACAGTCTGGGTGCTCGCCCACTTGT  
1710 1720 1730 1740 1750 1760 1770

1370 1380 1390 1400 1410 1420 1430  
TCTA--AGAGGAATTGTTGATACGTAGTATACTGTATAT-GTGTATG-TATATGTATATGTATA-TATAAGA  
| | | | | | | | | | | | | | | | | | | | | |  
CCCACTATCTGGGTGCATGATCTTGAGCAAGTTCCTTCTGGTGTCTGCTTTCTCCAT-TGTAAACCACAAGG  
1780 1790 1800 1810 1820 1830 1840 1850

1440 1450 1460 1470 1480 1490 1500  
CTCTTTTACTGTCAAAGTCAACCTAGAGTGTCTGGTTACCAGGTCAATTTTATTGG---ACATTTTACGTCA  
| | | | | | | | | | | | | | | | | | | | | |  
CTGTGCATGGGCTAA-TGA---AGA---TC---ATA-TACGTGAAAATTATTTGAAAACATATAAAG-CA  
1860 1870 1880 1890 1900



LOCUS HSIL05 6684 bp DNA PRI 03-JAN-1991  
 DEFINITION Human interleukin-2 (IL-2) gene and 5'-flanking region  
 ACCESSION X00695 X00200 X00201 X00202  
 KEYWORDS growth factor; interleukin; T-cell growth factor.  
 SOURCE human

ORGANISM Homo sapiens  
 Eukaryota; Animalia; Metazoa; Chordata; Vertebrata; Mammalia;  
 Theria; Eutheria; Primates; Haplorhini; Catarrhini; Hominidae.

REFERENCE 1 (bases 1 to 6684)

AUTHORS Holbrook,N.J., Lieber,M. and Crabtree,G.R.  
 TITLE DNA sequence of the 5' flanking region of the human interleukin 2  
 gene: homologies with adult T-cell leukemia virus  
 JOURNAL Nucleic Acids Res. 12, 5005-5013 (1984)  
 STANDARD full automatic

REFERENCE 2 (bases 1 to 6684)

AUTHORS Degraeve,W., Tavernier,J., Duerinck,F., Plaetinck,G., Devos,R. and  
 Fiers,W.  
 TITLE Cloning and structure of the human interleukin 2 chromosomal gene  
 JOURNAL EMBO J. 2, 2349-2353 (1983)  
 STANDARD full automatic

REFERENCE 3 (bases 1 to 6684)

AUTHORS Taniguchi,T., Matsui,H., Fujita,T., Takaoka,C., Kashima,N.,  
 Yoshimoto,R. and Hamuro,J.  
 TITLE Structure and expression of a cloned cDNA for human interleukin-2  
 JOURNAL Nature 302, 305-310 (1983)  
 STANDARD full automatic

FEATURES Location/Qualifiers  
 promoter 1339..1344  
 /note="TATA-box"  
 precursor\_RNA 1363..6403  
 /note="primary transcript"  
 intron 1563..1652  
 /note="intron I"  
 intron 1713..4004  
 /note="intron II"  
 intron 4149..6009  
 /note="intron III"  
 misc\_feature 6382..6387  
 /note="polyadenylation signal"  
 CDS join(1416..1562,1653..1712,4005..4148,6010..6117)  
 /product="interleukin"  
 /codon\_start=1  
 /translation="MYRMQLLSCLALSLALVTNSAPTSSSTKKTQLQLEHLLLDLQMI  
 LNGINNYKNPKLTRMLTFKPYMPKKATELKHLCLEELKPLEEVLNLAQSKNFHLRP  
 RDLISNINVIVLELKGSETTFMCEYADETATIVEFLNRWITFCQSIISTLT"

BASE COUNT 2342 a 1113 c 1064 g 2165 t  
 ORIGIN

Initial Score = 142 Optimized Score = 737 Significance = 7.26  
 Residue Identity = 47% Matches = 905 Mismatches = 732  
 Gaps = 257 Conservative Substitutions = 0

```

                                     X      10      20
                                     ATGTCATG--AACTGCTGAGT
                                     |  || |||  | ||   ||
ATGAATCACTTATTAGTGGACTGTTTCAGTTGAATTAATAAATACATTGAGATCAATGTCATCTAGACATT
 4880      4890      4900      4910      4920 X      4930      4940

      30      40      50      60      70      80
G-GATAAACAGCACGGGATATCTCTGTCTAAAG-----GAATATTACT-ACACCAGGAAAAGGACACAT
|  |  | |||  |  ||||| || |  ||      || || |  ||| ||| ||  || |||
GACAGATTCAAGTTC--CTTATCTATGGCAAGAGTTTTACTCTAAAAATAATTAACATCAGAAA----ACTCAT
 4950      4960      4970      4980      4990      5000

```

90 100 110 120 130 140 150

5010	5020	5030	5040	5050	5060	5070
------	------	------	------	------	------	------

5080            5090            5100            5110            5120            5130            5140

5150	5160	5170	5180	5190	5200	5210
------	------	------	------	------	------	------

5220                      5230                      5240                      5250                      5260                      5270

5280                      5290                      5300                      5310                      5320                      5330                      5340

5350            5360            5370            5380            5390            5400            5410

5420                      5430                      5440                      5450                      5460                      5470

5480            5490            5500            5510            5520            5530            5540

5550            5560            5570            5580            5590            5600            5610

5620            5630            5640            5650            5660            5670            5680

5690            5700            5710            5720            5730            5740            5750

020 070 040 050 060 070 080

ACCATTTAAGGAAGAC---CACTGGAGACGCTCAAGAGGAAGATGTTGTAGCTGCCGATGT--CCA-CAGGA  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
AAAATTTTGTGTTTCATTGACTGAATTAACAAATGAGGAAAT-CT--CAGCTTCTG-TGTTACTATCATTT  
5760 5770 5780 5790 5800 5810 5820

890 900 910 920 930 940 950  
AGAAGAAGGAGGAGGAGGAGGCTATGAGCTGTGATGTACTATCCTAGGAGATGTGTGGGCCGAAACCGAGAA  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
GGTATCATAACAA--AATACGCAAT----TTTGGCATTTC-AT-TTTGATCATTT----CAAGAAAATGTGAA  
5830 5840 5850 5860 5870 5880

960 970 980 990 1000 1010 1020  
GCAC TAGGACCCACCATCCTGTGGAACAGC---ACAAGCAACCCACACCTGTTCTTACACATCATCCT  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
TAATT-----AATATGTT-TGGTA-AGCTTGAAATAAAGGCAACAGGCC---TATAAGACTTCAATTG  
5890 5900 5910 5920 5930 5940

1030 1040 1050 1060 1070 1080 1090  
AGATGA--TGTGTGGGCGCGCACCTCATCCAAGTCTCTTCTAACGCTAACATATTTGTCTT-TACCTTTTTT  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
GGAATAACTGTATATAAGGTAACTACTC--TGTACTTTAAAAAATTAACATTTTTCTTTTATAGGGATCTG  
5950 5960 5970 5980 5990 6000 6010

1100 1110 1120 1130 1140 1150 1160  
AAATCTTTTTTAAATTTAAATTTTATG-TGTGTGAG--TGTTTTGCCTGCCTGTATGCACACGTGTGTGTG  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
AAA-CAACATTCATGTGTGAATATGCTGATGAGACAGCAACCATTG-TAGAATTTCTGAACAGATGGATTAC  
6020 6030 6040 6050 6060 6070 6080

1170 1180 1190 1200 1210 1220  
TGTGTGTGTGTGACACTCCTGATGCCTGAGGAGGTCA-GAAGAG--AAAGGGTTGGTTC-CATAAGAAC-TG  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
CTTTTGTCAAAGCATCATCTCAACACTGACTTGATAATTAAGTGCTTCCCACCTTAAACATATCAGGCCTTC  
6090 6100 6110 6120 6130 6140 6150

1230 1240 1250 1260 1270 1280 1290  
GAGTTATGGATGGCTGTGAGCCGNNNGATAGGTGGGACGGAGACCTGTCTTCTTATTTTAAC-GTGACTG  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
TATTTATTTAAATATTTAAATTTTATATTTATTGTTGAATGTATGGTTTGCTACCTATTGTAACATTTATTC  
6160 6170 6180 6190 6200 6210 6220 6230

1300 1310 1320 1330 1340 1350 1360 1370  
TATAATAAAAAAAAAATGATATTTGCGGAATTGTAGAGATT-GTCCTGACACCCCTTCTAGTTAATGATCTAA  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
T-TAATCTTAAACTAT-AAATAT-GGATCTTTTA-TGATTCTTTTGTAAAGCC--CTAG---GGGCTCTAA  
6240 6250 6260 6270 6280 6290

1380 1390 1400 1410 1420 1430  
GAGG-----AATTGTTGAT-ACGTAGTATACTGTA-TATG-TGTATGTATATGTATATGTATATAT-AAGAC  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
AATGGTTTCACTTATTTATCCCAAAATATTTATTATTATGTTGAATGT-TAAATATA-GTATCTATGTAGAT  
6300 6310 6320 6330 6340 6350 6360

1440 1450 1460 1470 1480 1490 1500  
TCTTTTACTGTCAAAGTCAACCTAGAGTGTCTGGTTACCAGGTCAATTTTATTGGACATT--TTACGTCACA  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
TGGTT---AGTAAAA--CTATTTAATAAATTTGATAA--ATATAACAAGCCTGGATATTTGTTATTTTGA  
6370 6380 6390 6400 6410 6420

1510 1520 1530 1540 1550 1560 1570  
CACA-CACACACACACACACACACGTTTATACTACGTAC---TGTTATCGGTATTCTACGTCATATAATG  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
AACAGCACAGA-GTAAGCATTTAAATATTTCT--TAGTTACTTGTGTGAAGTGTAGGAT-GGT--TAAAT-  
6430 6440 6450 6460 6470 6480 6490

1580 1590 1600 1610 1620 1630



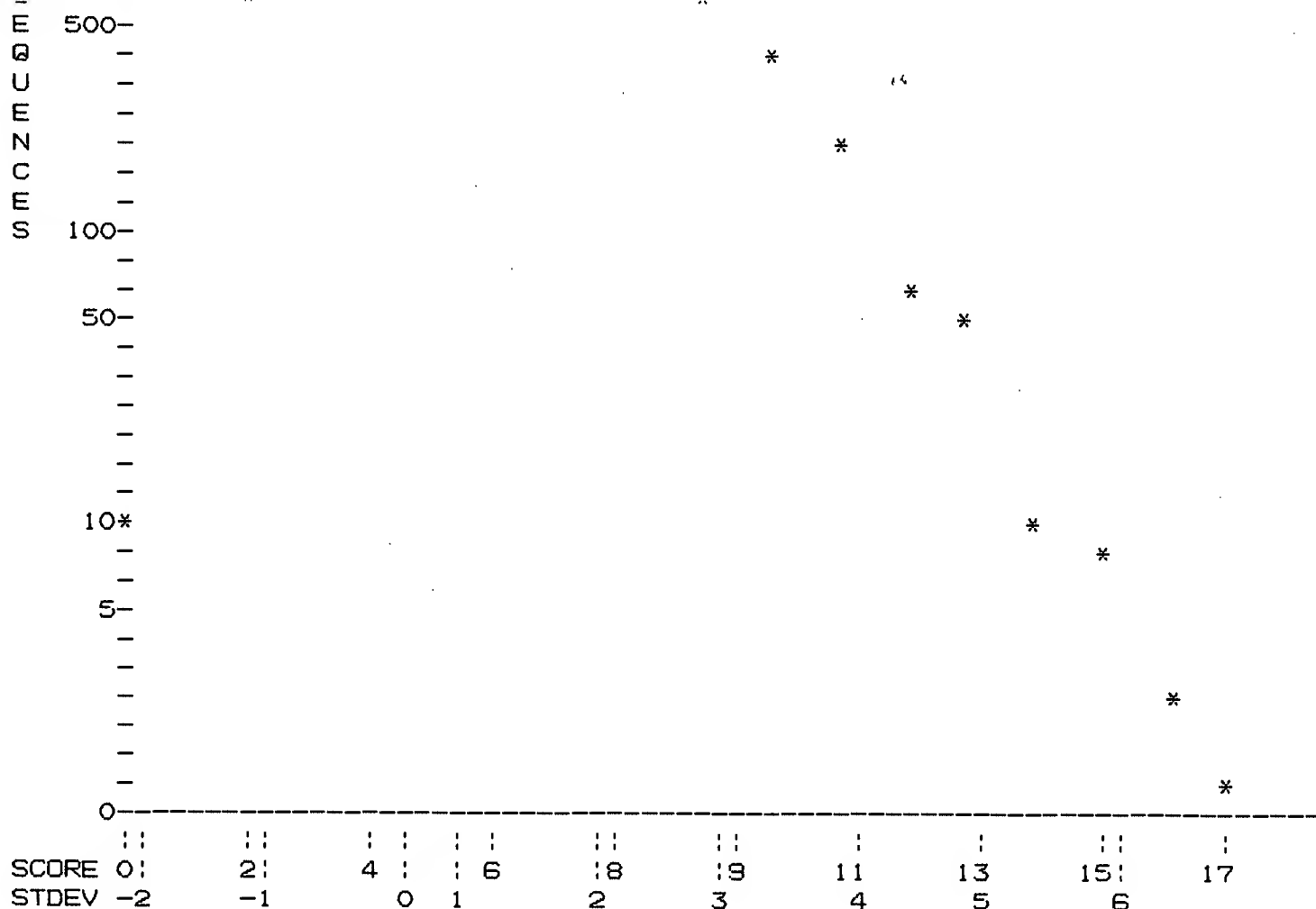


420            430            440            450            460            470            480  
CATTGCTTGGGGCCACAG--TGCACCAGATGTG-AAAAGGACTGCAGGC-CTGCCAGGAGCTAACGAAGCA  
|        |        |        |        |        |        |        |        |        |        |        |  
GAGCAC-CTCAGCTCTCTCTCAGCGGCACTGCAAGAAACACAGAGGCGCTCTTCCAGTCTTTCTGCAAGCA









#### PARAMETERS

Similarity matrix	Unitary	K-tuple	2
Mismatch penalty	1	Joining penalty	20
Gap penalty	1.00	Window size	32
Gap size penalty	0.05		
Cutoff score	5		
Randomization group	0		
Initial scores to save	20	Alignments to save	10
Optimized scores to save	20	Display context	10

#### SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	5	7	1.84

Times:	CPU	Total Elapsed
	00:02:34.02	00:07:35.00

Number of residues:	4627393
Number of sequences searched:	16524
Number of scores above cutoff:	4313

Cut-off raised to 6.  
 Cut-off raised to 7.  
 Cut-off raised to 8.

The scores below are sorted by initial score.  
 Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

4-1 BB

The list of best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
**** 6 standard deviations above mean ****						
1. A27381	Complement subcomponent C1s pr	688	17	43	6.51	0
**** 5 standard deviations above mean ****						
2. S01292	Tenascin - Chicken (fragment)	697	16	41	5.97	0
3. S01845	DNA (cytosine-5-)-methyltransf	1573	16	26	5.97	0
4. VVVPBD	Coat protein VP1 - Budgerigar	343	15	30	5.43	0
5. PS0047	Extracellular serine protease	448	15	44	5.43	0
6. A27733	nifA protein - Azotobacter vin	129	15	25	5.43	0
7. S01927	Regulatory protein nifA - Azot	522	15	38	5.43	0
8. W6WLHS	Probable E6 protein - Papillom	158	15	28	5.43	0
9. S04029	Sodium channel protein - Fruit	1321	15	35	5.43	0
10. D31090	Hydrogen ion-transporting ATP	163	15	22	5.43	0
**** 4 standard deviations above mean ****						
11. MNXRW4	Nonstructural protein Pns4 - W	732	14	44	4.89	0
12. ZLVN	L protein - Vesicular stomatit	2109	14	43	4.89	0
13. B28392	Penicillin amidase I precursor	558	14	41	4.89	0
14. DEECDA	Aspartate-semialdehyde dehydro	367	14	25	4.89	0
15. WMBEH6	UL36 protein - Herpes simplex	3164	14	36	4.89	0
16. S01165	Achaete-scute locus protein T3	257	14	37	4.89	0
17. KXB0Z	Protein Z - Bovine	396	14	34	4.89	0
18. VCLJB	env polyprotein - Bovine leuke	515	14	44	4.89	0
19. S06053	Transforming protein (ski) - H	728	14	40	4.89	0
20. QQBE6L	Hypothetical BXL2 protein - E	706	13	47	4.34	0

The scores below are sorted by optimized score.

Significance is calculated based on optimized score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
**** 5 standard deviations above mean ****						
1. VGVUPT	Glycoprotein precursor - Punta	1313	9	51	5.18	0
2. EGMSMG	Epidermal growth factor precu	1217	9	51	5.18	0
**** 4 standard deviations above mean ****						
3. TVRTNU	Kinase-related transforming pr	1260	9	50	4.71	0
4. A30359	Granule membrane protein 140 p	830	7	50	4.71	0
5. AHRB	Ig alpha chain C region - Rabb	299	9	50	4.71	0
6. KQHUP	Plasma kallikrein precursor -	638	8	50	4.71	0
7. GQHUN	Nerve growth factor receptor p	427	8	49	4.24	0
8. QRHULD	LDL receptor precursor - Human	860	7	49	4.24	0
9. W2WLB2	Probable E2 protein - Bovine p	422	7	49	4.24	0
10. JL0104	Lymphocyte-associated cell sur	385	9	49	4.24	0
11. A26850	Hydrogen ion-transporting ATP	489	10	49	4.24	0
12. VHWVB	Structural polyprotein - Sindb	1245	13	49	4.24	0
13. S06028	Gene supressor-of-white-aprico	964	7	49	4.24	0
**** 3 standard deviations above mean ****						
14. A28455	Cell surface antigen 4F2 heavy	529	8	48	3.77	0
15. A32375	Lymphocyte surface MEL-14 anti	372	8	48	3.77	0
16. GNVUUK	Glycoprotein precursor - Uukun	1008	11	48	3.77	0
17. SYECCP	Carbamoyl-phosphate synthase (	1072	8	48	3.77	0
18. MHMS	Ig mu chain C region - Mouse	455	8	48	3.77	0
19. UIBO	Thyroglobulin precursor - Bovi	2769	8	48	3.77	0
20. A24976	Ig mu chain C region, b allele	455	8	48	3.77	0

## 1. ELLIS-267-3A

VGUPT

Glycoprotein precursor - Punta Toro virus

ENTRY VGUPT #Type Protein  
 TITLE Glycoprotein precursor - Punta Toro virus  
 INCLUDES glycoprotein NS-M\ glycoprotein G1\ glycoprotein G2  
 DATE 27-Nov-1985 #Sequence 27-Nov-1985 #Text 31-Dec-1989  
 PLACEMENT 1707.0 1.0 1.0 1.0 1.0  
 SOURCE Punta Toro virus  
 ACCESSION A04109  
 HOST #Common-name mosquito\  
 Homo sapiens #Common-name man  
 REFERENCE (Sequence translated from the RNA sequence)  
 #Authors Ihara T., Smith J., Dalrymple J.M., Bishop D.H.L.  
 #Journal Virology (1985) 144:246-259  
 COMMENT This virus is a member of the family Bunyaviridae.  
 SUPERFAMILY #Name phlebovirus glycoprotein  
 KEYWORDS glycoprotein\ transmembrane protein  
 FEATURE  
 1-270 #Protein glycoprotein SN-M (SNM)\  
 271-809 #Protein glycoprotein G1 (GG1)\  
 810-1313 #Protein glycoprotein G2 (GG2)\  
 76,102,1,1243 #Binding-site carbohydrate (possible)  
 SUMMARY #Molecular-weight 146374 #Length 1313 #Checksum 4967  
 SEQUENCE

Initial Score = 9 Optimized Score = 51 Significance = 5.18  
 Residue Identity = 23% Matches = 68 Mismatches = 172  
 Gaps = 53 Conservative Substitutions = 0

```

      X      10      20      30      40      50
      MGNNCYNVVVIVLLLVGCEKVGAVQNSCDNCQPG--TFCKRY-----NPVCKSCPPSTFS
      |      |      |      |      |      |      |      |      |      |
TNVSFVCYEHVGQDEQVEVHRALKRVS VNDCKIVDNSKQKICTGDHVFCEKYDCSTSYPDVTCIHAPGSGPL
      500      X 510      520      530      540      550      560

      60      70      80      90      100      110
      SIGGQPNCNICRVCAGY-----FRFKKFCSSSTHNAEC-----EC-IEGFHCLGPQCTRCEKD CRPGQELT-K
      |      |      |      |      |      |      |      |      |      |
YI-NLMGSWIKPQCVGYERVLVDREVKQPLL APEQNC DTCVSECLDEGVH-----IKSTGFEITSA
      570      580      590      600      610      620

      120      130      140      150      160      170      180
      QGCKTCSLGT FNDQNGT--GVC RPWTNCSLDGRSVLKTGTTEKDVVCGPPVVSFSP--STTISVTPEGGPGG
      |      |      |      |      |      |      |      |      |      |
VACSHGSCISAHQEPSTSVIVPYPGLLASVGG RIGIHL SHT-SDSASVHMVVVCPPRDSCA AHNCLLCYHGI
      630      640      650      660      670      680      690

      190      200      210      220      230      240
      HSLQ-VLTLFLALTSALLLALIFITLLFSVLKWIRKKFPHIFKQPFKKTG-----AAQEEDACSC
      |      |      |      |      |      |      |      |      |      |
LNYQCHSTLSAILTSFLL--ILFIYTVFSVTTN ILYVLR LIPKQ-LKSPVGWLKLFINWLLTALRIKTRNVM
      700      710      720      730      740      750      760

      250      X
      RCPQEEEGGGGGYEL
      |      |
      RRINQ RIGVVDHHDVERPRHREPMR
      770      780      790

```

## 2. ELLIS-267-3A

EGMSMG

Epidermal growth factor precursor - Mouse

ENTRY EGMSMG #Type Protein

TITLE Epidermal growth factor precursor - Mouse  
 DATE 30-Nov-1980 #Sequence 11-Aug-1983 #Text 31-Dec-1989  
 PLACEMENT 575.0 1.0 1.0 1.0 1.0  
 SOURCE Mus musculus #Common-name house mouse  
 ACCESSION A01387  
 REFERENCE (Sequence translated from the mRNA sequence)  
 #Authors Scott J. , Urdea M. , Quiroga M. , Sanchez-Pescador R. ,  
 Fong N. , Selby M. , Rutter W. J. , Bell G. I.  
 #Journal Science (1983) 221:236-240  
 #Comment The cleavage site for the signal sequence is not  
 known.  
 #Comment The precursor sequence contains seven regions that  
 are similar to the epidermal growth factor  
 sequence: residues 357-399, 400-440, 441-480,  
 745-784, 832-885, 886-925, and 926-976.  
 REFERENCE (Sequence of residues 1-1168 translated from the  
 mRNA sequence)  
 #Authors Gray A. , Du11 T. , Ullrich A.  
 #Journal Nature (1983) 303:722-725  
 #Comment This sequence differs from residues 1-1133 of that  
 shown in having 790-Tyr and 1048-Ser. It differs  
 greatly from residues 1134-1168 of that shown due  
 to an insertion of one base in the nucleotide  
 sequence with respect to the nucleotide sequence  
 of Scott, et al. , which causes a shift in the  
 reading frame.  
 #Comment There are sequence homologies between residues  
 321-360, 361-401, 402-442, 443-482, 746-786,  
 837-875, 876-917, 918-958, and 978-1018.  
 REFERENCE (Active protein, complete sequence of residues  
 977-1029 with experimental details)  
 #Authors Savage Jr C. R. , Inagami T. , Cohen S.  
 #Journal J. Biol. Chem. (1972) 247:7612-7621  
 #Comment Residues 1024-1029 are not required for full  
 biological activity in vivo.  
 REFERENCE (Disulfide bonds)  
 #Authors Savage Jr C. R. , Hash J. H. , Cohen S.  
 #Journal J. Biol. Chem. (1973) 248:7669-7672  
 #Comment Disulfide bonds link residues 982-996, 990-1007, and  
 1009-1018.  
 COMMENT The active growth factor from this submaxillary  
 gland protein stimulates the growth of various  
 epidermal and epithelial tissues in vivo and in  
 vitro and of some fibroblasts in cell culture.  
 SUPERFAMILY #Name epidermal growth factor  
 SUMMARY #Molecular-weight 133143 #Length 1217 #Checksum 9280  
 SEQUENCE

Initial Score = 9 Optimized Score = 51 Significance = 5.18  
 Residue Identity = 23% Matches = 69 Mismatches = 162  
 Gaps = 60 Conservative Substitutions = 0

```

      X      10      20      30      40      50      60
      MGNNCYNVVVIVLLLLVGCEKVGAVQNSCDNCQPGTFCRKYNPVCKSCPPSTFFSIGGQPNCN
      ::      :      :      :      :      :      :
SCFDIDECQRGAHNCAENAACTNTEGGYNCTCAGRPS----SPGRSC-----PDSTAPSLLGEDGHHLDRN
    920      X 930      940      950      960      970

      70      80      90      100      110
      ICRVC----AGYFRFKKFC---SSTHNAECECIEGFHCLGPQCTRCEKDCR-----PGQELTKQGCKTC
      :      ::      :      :      :      :      :      :
      SYPGCPSSYDGYCLNGGVCMHIESLDSYTCNCVIGYSGDRCQ-TR---DLRWELRHAGYGQKHDIMVAVC
    980      990      1000      1010      1020      1030      1040

    120      130      140      150      160      170
    ----SLGTFNDQNGTGVCR-----PWTN-CSLDGRSVLKTGTTEKDVVCGPPVVSFSPSTTISVTPEGGP
  
```



MVALVLLLLLGMWGTYYYRTRKPLSNPPKNPCDEPSGSVSSSG--PDSSSGAAVASCPQPFVVLKHKQDP  
1050 1060 1070 1080 1090 1100 1110

```

180      190      200      210      220      230      240
GGHSLQVLTLFLALTSALLLALIFITLLFSV-LKWIRKKFPHIFKQPFKKTGAAGQEDACSRCPQEEEG-
      ::      :      :      :      :      :      :      :      :      :      :
KNGSLPADGTNGAVVDA--GLSPSLQLGSHLTWRQK-PHI---DGMGTGQSCWIPPSSDRGPQEIEN
      1120      1130      1140      1150      1160      1170

      250      X
-----GGGGYEL
      :      :
SHLPSYRPVGPEKLHSLQSANGS
1180      1190      1200

```

3. ELLIS-267-3A

TVRTNU	Kinase-related transforming protein precursor (neu
--------	--

```
ENTRY          TVRTNU          #Type Protein
TITLE          Kinase-related transforming protein precursor (neu)
                - Rat #EC-number 2.7.1.-
DATE          31-Dec-1988 #Sequence 31-Dec-1988 #Text 31-Dec-1988
PLACEMENT      197.0    15.0    2.0    1.0    2.0
SOURCE         Rattus norvegicus #Common-name Norway rat
ACCESSION      A24562
REFERENCE      (Sequence translated from the mRNA sequence)
                #Authors  Bargmann C. I. , Hung M. C. , Weinberg R. A.
                #Journal  Nature (1986) 319:226-230
                #Title    The neu oncogene encodes an epidermal growth factor
                        receptor-related protein.
```

```

GENETIC      #Name          neu
SUPERFAMILY  #Name kinase-related transforming protein
KEYWORDS     transforming protein\ tyrosine-specific protein
              kinase

```

```

FEATURE
  1-19          #Domain signal sequence (SIG)\
  20-1260       #Protein kinase-related transforming
                protein neu (KTP)\
  658-680       #Domain transmembrane (TMN)\
  731-986       #Domain tyrosine-specific protein kinase
                (TPK)\

  71,191,263,535,576,
  634,763,1146,1231    #Binding-site carbohydrate (Asn)
                        (possible)\
  691,882,1227,1253    #Modified-site phosphorylation

SUMMARY      #Molecular-weight 139219  #Length 1260  #Checksum 5917
SEQUENCE

```

Initial Score	=	9	Optimized Score	=	50	Significance	=	4.71
Residue Identity	=	23%	Matches	=	69	Mismatches	=	158
Gaps	=	67	Conservative Substitutions	=	0			

[illegible]

Best Available Copy

```

110      120      130      140      150      160      170
GGELTKGGCK--TCSLGTFNDGNTGVCRFWYNCSLDGRSVLKTGTTEKDVVCGPPVVSFSPSTTISVTPEG
      | |      | | |      |      |      |      |      |      |
CVARCPSGVKPDLSYMPIWKYPDEEGICQPCPINCTHSCVDL-----DERGCPAEQRASPVTFIIATVEG
      610      620      630      640      650      660

180      190      200      210      220      230      240
GPGGHSLQVLTLFLALTSALLLALIFITLLFSVLKWIRKKFPHIFKQ----PFKKTTGAAGQEDACSCRCPO
      | | | | |      | |      | | |      |      |      |      |
-----VL-LFLIL-VVVVGILI-----KRRRQKIRKYTMRRLLQETELVEPLTPSGAMPNQA-QMRILK
      670      680      690      700      710      720

      250      X
EEE-----GGGGGYEL
      | |      | |
ETELRKVKVLGSGAFGTVYKGIWIPD
      730      740

```

#### 4. ELLIS-267-3A

A30359 Granule membrane protein 140 precursor - Human

ENTRY A30359 #Type Protein  
 TITLE Granule membrane protein 140 precursor - Human  
 SOURCE Homo sapiens #Common-name man  
 ACCESSION A30359  
 REFERENCE (Sequence translated from the mRNA sequence)  
 #Authors Johnston G. I., Cook R. G., McEver R. P.  
 #Journal Cell (1989) 56:1033-1044  
 #Title Cloning of GMP-140, a granule membrane protein of platelets and endothelium: sequence similarity to proteins involved in cell adhesion and inflammation.

FEATURE  
 1-41 #Domain signal sequence (SIG)\  
 42-830 #Protein granule membrane protein 140 (MAT)\  
 42-159 #Domain lectin (LEC)\  
 160-199 #Domain EGF (EGF)\  
 772-795 #Domain transmembrane (TMN)\  
 54,98,180,212,219,411,  
 460,518,665,716,723,  
 741 #Binding-site carbohydrate (Asn)\  
 200-770 #Domain complement H/C4b-binding (COM)

COMMENT THIS SEQUENCE HAS NOT BEEN COMPARED TO THE NUCLEOTIDE TRANSLATION.

SUMMARY #Molecular-weight 90766 #Length 830 #Checksum 2552  
 SEQUENCE

Initial Score = 7 Optimized Score = 50 Significance = 4.71  
 Residue Identity = 22% Matches = 66 Mismatches = 174  
 Gaps = 47 Conservative Substitutions = 0

```

      X      10      20      30      40      50      60
      MGNNCYNVVVIVLLLLVGCEKVGAVQNSCDNCQPGTFCRKYNPVCKSCPPSTFSSIGGQPNCN
      | | | | | | | | | | | | | | | | | | | | | |
NEARVNC SHPFGAFRYQSVCSFTCNEGLLLVGA---SVLQCLATGNWNSVPPECBAIPCTPLLS--PQNGTM
      460      X      470      480      490      500      510      520

      70      80      90      100      110      120
ICRV CAGYFRFKKFCSSTHNAECECIEGFHCLGP---QCTR--CEKDCRPGQELTK-----QGCKTCS-
      | | | | | | | | | | | | | | | | | | | | | |
TCVQPLGSSSYKSIC-----QFICDEGYSLSGPERLDCTRSGRWTDSPPMCEAIKCPELFAPEQGGSLDCSD
      530      540      550      560      570      580

```

Best Available Copy

```

130      140      150      160      170      180
-LGTFN-----DNGTGVCRPHTNGSLDGRSVLKTGTTEKDVVCGPPVVSFSPSTTISVTPEGGPGGH
:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
TRGEFNVGSTCHFSCNNGFKLEGPNNVECTTSGR-WSATPPTCKGIASLPTPGLQCPALT---TPGQGTMYC
590      600      610      620      630      640      650

      190      200      210      220      230      240
SLQVLTLFLALTSALLLALIFITLLFSVLKWIRKKFPHIFKQP----FKKTTGAAQEEEDA--CSCRCPEEEE
:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
RHHPGT-FGFNTTCYFGCNAGFTLIGDSTLSCRPSGQWTAVTPACRAVKCSELHVNKPIAMNCSNLWGNFSY
660      670      680      690      700      710      720

250      X
G-GGGGYEL
:  :  :
GSICSGHCLEGQLLNGSAQ
730      X  740

```

# 5. ELLIS-267-3A

AHRB Ig alpha chain C region - Rabbit (fragment)

```

ENTRY      AHRB      #Type Protein (fragment)
TITLE      Ig alpha chain C region - Rabbit (fragment)
DATE       28-Aug-1985 #Sequence 28-Aug-1985 #Text 30-Jun-1989
PLACEMENT  696.0   13.0   3.0   1.0   1.0
SOURCE     Oryctolagus cuniculus #Common-name domestic rabbit
ACCESSION  A02174
REFERENCE  (Sequence translated from the mRNA sequence)
#Authors   Knight K. L. , Martens C. L. , Stoklosa C. M. ,
            Schneiderman R. D.
#Journal   Nucleic Acids Res. (1984) 12:1657-1670
COMMENT    This immunoglobulin belongs to the IgA-g subclass.
            It was isolated from a rabbit homozygous for a2,
            n80, del12,15, f71, g75 heavy chain haplotype.
SUPERFAMILY #Name immunoglobulin C region
KEYWORDS    immunoglobulin\ plasma protein
SUMMARY
SEQUENCE    #Length 299 #Checksum 2361

```

```

Initial Score      =      9   Optimized Score      =      50   Significance      =      4.71
Residue Identity   =     23%   Matches              =      69   Mismatches       =     153
Gaps               =      70   Conservative Substitutions      =      0

```

```

X      10      20      30      40      50      60
MGNNCYNVVIVLLLVGCEKVGAVQNSCDNQPGTFCKYNPVCKSCPPSTFSSIGGQPNC---NICRVCA
:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QSGTSGPYTACSELILPVTQCLG--QKS-AAC-----HVEYNSVINESLPVPF-----PDCCPANSCTC-
X      10      20      30      40      50

70      80      90      100      110      120      130
GYFRFKKFCSSSTHNAECECIEGFHCLGPQCTRCEKDCRPGQELTKQGCK--TCSLGTFTNDQNGTGVCRPWTN
:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
-----PSSSSRNLI SGCQPSLSLQRPDLGDL L LGRDASLTCTLSGLKNPEDAVFTWEPTNGNEPVQQRAG
      60      70      80      90      100      110      120

140      150      160      170      180      190
CSLDG----RSVL-----KTGTTEKDVVCGPPVVSFSPSTTIS---VTPEGGPGGHS LQVLTLFLAL TSA
:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
RDLSGCYSVSSVLPSSAETWKARTEFTCTVTHPEIDSGSLTATISRGVVTP---PQVHLLPPPSEELALNEQ
      130      140      150      160      170      180      190

200      210      220      230      240      250
LLLALIFITLL---FS---VLKWIR-----KKFPHIFKQP---FKKTTGAAQEEEDACSCRCPEEEGGG
:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
VTL-----TCLVRGFSPKDVLSWRHQGGQEVPEDSFLVWKSMPESQDKATYA----ITSLLRVPADWNQG

```

X  
GGYEL  
:  
DTYSCMVGHEGLAEH  
260

# 6. ELLIS-267-3A

KQHUP Plasma kallikrein precursor - Human #EC-number

ENTRY KQHUP #Type Protein  
TITLE Plasma kallikrein precursor - Human #EC-number  
3. 4. 21. 34  
ALTERNATE-NAME plasma prekallikrein\ kininogenin  
DATE 13-Aug-1986 #Sequence 13-Aug-1986 #Text 13-Aug-1986  
PLACEMENT 356.0 4.0 2.0 1.0 1.0  
SOURCE Homo sapiens #Common-name man  
ACCESSION A00921  
REFERENCE (Sequence translated from the mRNA sequence)  
#Authors Chung D. W. , Fujikawa K. , McMullen B. A. , Davie E. W.  
#Journal Biochemistry (1986) 25:2410-2417  
COMMENT This protein, synthesized in the liver, circulates  
as a noncovalent complex with high molecular  
weight (HMW) kininogen.  
COMMENT The zymogen is activated by factor XIIa, which  
cleaves the molecule into a light chain, which  
contains the active site, and a heavy chain, which  
associates with HMW kininogen. These chains are  
linked by one or more disulfide bonds.  
COMMENT The enzyme cleaves Lys-Arg and Arg-Ser bonds. It  
activates, in a reciprocal reaction, factor XII  
after its binding to a negatively charged surface.  
It also releases bradykinin from HMW kininogen and  
may also play a role in the renin-angiotensin  
system by converting prorenin into renin.  
SUPERFAMILY 389-621 #Name trypsin  
KEYWORDS hydrolase\ serine proteinase\ glycoprotein\ plasma\  
blood coagulation\ fibrinolysis\ inflammation\  
liver\ duplication  
FEATURE  
1-19 #Domain signal sequence (SIG)\  
20-390,391-638 #Protein plasma kallikrein, heavy and  
light chains (MPT)\  
389-621 #Domain (or 383-625) serine proteinase  
(TRY)\  
20-104,110-194,  
200-284,291-375 #Duplication\  
434 #Active-site His\  
483 #Active-site Asp\  
578 #Active-site Ser\  
127,308,396,453,494 #Binding-site carbohydrate (Asn)  
SUMMARY #Molecular-weight 71369 #Length 638 #Checksum 585  
SEQUENCE

Initial Score = 8 Optimized Score = 50 Significance = 4.71  
Residue Identity = 23% Matches = 70 Mismatches = 164  
Gaps = 69 Conservative Substitutions = 0

X 10 20 30 40 50  
MGNNC-----YNVVVI-----VLLLVGCEKVGAVQNSCDNCQPGTFCKYNPVCKSCPST  
:: :: :: :: :: :: :: :: :: :: :: ::  
DAFVCRTICTYHPNCLFFTFYTNVWKIESQBNVCLLKTSE-SGTPSSS--TPQENTISGYSLLTCKRTLPEP  
230 X 240 250 260 270 280 290

Best Available Copy

```

FSS---IGGQPN CNICRV CAGYFRFK---FCSSTHNAECECIEGFHCLGPQCTRCEKDCRPGQELTKQGCK
|         |         |         |         |         |         |         |         |
CHSKIYPGVDFGGEELNV-----TFVKGVNVCQETCTKMIRCQFFTYSLLPEDCKEEK-CKCFLRLSMDGSP
      300           310           320           330           340           350

      120           130           140           150           160           170
T-----CSLGTFNDAQNGTG---VCRPWTNCSLDGRSVLKTGTTEKDVVCGPPVVSFSPSTTISVTPEGGP
|         |         |         |         |         |         |         |         |
TRIAYGTQGGSSGYSLRLCNTGDNSVCTTKT----STRIV---GGTNSSWGEWPWQVSLQVKLTAGRHLCCGGS
360           370           380           390           400           410           420

180           190           200           210           220           230
GGHSLQVLT-----LFL----ALTSALL-LALIFITLLFSVLKWIRKKFPHIFKQPFKKTG-----AAQE
|         |         |         |         |         |         |         |         |
LIGHQWVLTAAHCFDGLPLQDVWRIYSGILNLSDITKDTPFSSQIKEI-----IIHQNYKVSEGNHDIALIK
      430           440           450           460           470           480

      240           250           X
EDA----CSCRCPEEEEEGGGGGYEL
|         |         |         |
LQAPLNYTEFQKPICLPSKGGDTSTIYTNCWVTGWG
490           500           510 X           520

```

# 7. ELLIS-267-3A

GQHUN Nerve growth factor receptor precursor - Human

```

ENTRY      GQHUN      #Type Protein
TITLE      Nerve growth factor receptor precursor - Human
ALTERNATE-NAME NGF receptor
DATE       31-Mar-1988 #Sequence 31-Mar-1988 #Text 31-Mar-1988
PLACEMENT  580.0      1.0      1.0      1.0      1.0
SOURCE     Homo sapiens #Common-name man
ACCESSION  A25218
REFERENCE  (Sequence translated from the mRNA sequence)
  #Authors Johnson D. , Lanahan A. , Buck C.R. , Sehgal A. , Morgan
            C. , Mercer E. , Bothwell M. , Chao M.
  #Journal Cell (1986) 47:545-554
  #Title    Expression and structure of the human NGF receptor.
COMMENT    This receptor is found on sensory and sympathetic
            neurons, on neuroblastoma cells, and on a variety
            of nonneuronal derivatives of the neural crest.
COMMENT    The duplicated cysteine-rich region of the
            extracellular domain may form part or all of the
            NGF-binding site. The active form of NGF is a
            noncovalent dimer of identical chains.
COMMENT    Although structurally similar, this receptor differs
            from other growth factor receptors in that its
            cytoplasmic domain is not homologous to known
            tyrosine or serine/threonine protein kinases.
            Although apparently lacking intrinsic kinase
            activity, it is phosphorylated on serine.
COMMENT    This receptor undergoes both N- and O-linked
            glycosylation.
GENETIC
  #Map-position 17q21-q22
  #Name         NGFR
SUPERFAMILY  #Name nerve growth factor receptor
KEYWORDS     receptor\ integral membrane protein\ glycoprotein\
            duplication
FEATURE
  1-28        #Domain signal sequence <SIG>\
  29-427      #Protein nerve growth factor receptor
            <MAT>\
  29-250      #Domain extracellular <EXT>\

```

25-100/105-130  
 29-190  
 197-248  
 251-272  
 273-427  
 60

Best Available Copy

#Duplication\  
 #Region cysteine-rich\  
 #Region serine/threonine-rich\  
 #Domain transmembrane (MEM)\  
 #Domain cytoplasmic (CYT)\  
 #Binding-site carbohydrate (Asn)  
 (putative)

SUMMARY #Molecular-weight 45183 #Length 427 #Checksum 7426  
 SEQUENCE

Initial Score = 8 Optimized Score = 49 Significance = 4.24  
 Residue Identity = 22% Matches = 66 Mismatches = 173  
 Gaps = 54 Conservative Substitutions = 0

```

      X      10      20      30      40      50
      MGNNCYNVVVIVLLLVGCEKVGAVG-----NSCDNCQPG-----TFCRKYNPVCKSCPPS
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
PCTECVGLQSMSAPC----VEADDAVCRCAVGYYQDETTGRCEACRVCEAGSGLVFSCQDKQNTVCEECPDG
      90      X      100      110      120      130      140      150

      60      70      80      90      100      110
      TFS-SIGGGPNCNICRVCAGYFRFKKFCSSSTHNAECECIEG-----FHCLGPQCT---RCEKDCRPGQEL-
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
TYSDEANHVDPCLPCTVCEDTERQLRECTRWADAECEEIPGRWITRSTPPEGSDSTAPSTQEPPEAPPEQDLI
      160      170      180      190      200      210      220

      120      130      140      150      160      170      180
      --TKQGCKTCSLGTENDQNGTGVCRPWT-NCSLDGRSVLKTGTTEKDVVCGPPVVSFSPSTTISVTPEGGPG
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
ASTVAGVVT----TVMGSSQPVVTRGTTDNLIPVYCSIL-----AAVVVG-LVAYIAFKRWNS-CKQNKQG
      230      240      250      260      270      280

      190      200      210      220      230      240
      GHSLQVLTLFLALTSALLLALIFITLLFSVLKWIRKKKPFHIFKQPFKKTTG-----AAQEE----DACSC
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
ANSRPV-NQTPPPEGEKLSHDSGISVDSQSLHDQQPHQTATASGQALKGDGGLYSSLPPAKREEVEKLLNGSA
      290      300      310      320      330      340      350

      250      X
      RCPQEEEGGGGGYEL
      | |
      GDTWRHLAGELGYQPEHIDSFTHEA
      360      370      380

```

# 8. ELLIS-267-3A

QRHULD LDL receptor precursor - Human

ENTRY QRHULD #Type Protein  
 TITLE LDL receptor precursor - Human  
 DATE 17-May-1985 #Sequence 17-May-1985 #Text 28-May-1986  
 PLACEMENT 574.0 1.0 1.0 1.0 1.0  
 SOURCE Homo sapiens #Common-name man  
 ACCESSION A01383  
 REFERENCE (Sequence translated from the mRNA sequence)  
 #Authors Yamamoto T., Davis C.G., Brown M.S., Schneider W.J.,  
 Casey M.L., Goldstein J.L., Russell D.W.  
 #Journal Cell (1984) 39:27-38  
 COMMENT This transmembrane glycoprotein binds LDL, the major  
 cholesterol-carrying lipoprotein of human plasma,  
 and transports it into cells by endocytosis. In  
 order to be internalized, the receptor-ligand  
 complexes must first cluster into clathrin-coated  
 pits.  
 COMMENT The amino end of the extracellular domain contains  
 seven or eight 40-residue repeats. Each repeat has

involved in disulfide bonds. Following these repeats is a region of about 350 residues that is homologous with part of the epidermal growth factor (EGF) precursor.

COMMENT The last half of the extracellular domain contains structural evidence of repetitive sequence in the similarity of residues 441-445, 488-492, 531-535, 575-579, and 617-621.

COMMENT An intrastrand recombination event between two Alu sequences in the 3' untranslated region of mRNA from a familial hypercholesterolemia patient results in the deletion of the transmembrane and cytoplasmic domains. Most of the receptors produced are secreted, but those that adhere to the cell surface cannot cluster in coated pits; therefore, even though they bind LDL, these receptor-ligand complexes are not internalized.

SUPERFAMILY #Name LDL receptor

KEYWORDS glycoprotein\ LDL\ cholesterol\ lipid transport\ endocytosis\ coated pits\ transmembrane protein\ receptor

#### FEATURE

22-860 #Protein LDL receptor (MAT)\  
1-21 #Domain signal sequence (SIG)\  
22-788 #Domain extracellular (EX1)\  
22-61,62-102,103-141,  
142-180,191-229,  
230-268,269-309 #Duplication\  
311-661 #Region EGF precursor homology\  
721-768 #Region clustered O-linked  
oligosaccharides\  
789-810 #Domain transmembrane (TMM)\  
811-860 #Domain cytoplasmic (CYT)

SUMMARY #Molecular-weight 95375 #Length 860 #Checksum 3641

#### SEQUENCE

Initial Score = 7 Optimized Score = 49 Significance = 4.24  
Residue Identity = 23% Matches = 67 Mismatches = 162  
Gaps = 56 Conservative Substitutions = 0

```

X      10      20      30      40      50      60
MGNNCYNVVVIVLLL VGCEKVGAVQNSC----DNCQPGTFCRKYNPVCKSCPPSTFSSIGGQPN CNICRVCA
::      :  ::      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
MGPWGWLRLRTVALL-LAAAGTAVGDR CERNEFQCQDG-KCISYK WVC DGS AECQDGSDESQETCLSVTCKS
X      10      20      30      40      50      60      70
70      80      90      100     110     120     130
GYFRFKKFCSSSTHNAECECIEGFHCLGPQCTRC--EKDCRP GQELTKQGC--KTCSLGT FNDQNGTGVC RPW
:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
GDF----SCGGRVN---RCI-----PQFWRC D GQVDCDNG--SDEQGC PPKTCSQDEF RCHD GKCISRQF
      80      90      100     110     120
140      150     160     170     180     190
TNCS----LDG-----RSVLKTGTTEKDVVCGPPVVSFSPSTTI-----SVTPEGGPGGHSLQVLTL-F
:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
VCDSDRDCLDGSDEASCPVL-----TCGPASFQCNSSTCIPQLWACDNDPDCEDGSDEWPQRCRGLYV
130      140      150     160     170     180
200      210     220     230     240     250     X
LALTSALLLALIFITL--LFSVLKWIRKKFP-HIFKQPFKKT TGA AQEEDACSCRC PQEEEGGGGGYEL
:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
FQGDSSPCSAFEFHCLSGECIHSSWRC D GGP DCKDKSDEENCAVATCRPDEFQCS D GNCIHGSRQCDREYDC
190      200     210     220     230     240     250     260

```

## 9. ELLIS-267-3A

W2WLB2 Probable E2 protein - Bovine papillomavirus (type

ENTRY W2WLB2 #Type Protein  
 TITLE Probable E2 protein - Bovine papillomavirus (type 2)  
 DATE 31-Mar-1989 #Sequence 31-Mar-1989 #Text 31-Mar-1989  
 PLACEMENT 1269.0 7.0 1.0 2.0 1.0  
 SOURCE bovine papillomavirus  
 ACCESSION D31169  
 REFERENCE (Sequence translated from the DNA sequence)  
 #Authors Groff D. E. , Mitra R. , Lancaster W. D.  
 #Citation submitted to GenBank, May 1988  
 COMMENT The DNA sequence was obtained from GenBank, release 57.0.  
 COMMENT This virus is a member of the family Papovaviridae.  
 SUPERFAMILY #Name papillomavirus E2 protein  
 KEYWORDS early protein  
 SUMMARY #Molecular-weight 46877 #Length 422 #Checksum 6025  
 SEQUENCE

Initial Score = 7 Optimized Score = 49 Significance = 4.24  
 Residue Identity = 23% Matches = 66 Mismatches = 169  
 Gaps = 50 Conservative Substitutions = 0

```

      X          10          20          30          40          50
      MGN--NCYNVVVIVLLLLVGCE-----KVGAVQNSCDNCQPGTFCRKYNPVCKSCPPSTFSS
      ::  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
    KGARVVEVEFDGNASNTNWTYTVYSKLYMRTEDGWQLAKAGADGTGLYYCTMAGAGRIY-YSRFGEEAARFST
      130 X   140       150       160       170       180       190

      60          70          80          90          100          110          120
    IGGQPNICRVCAGYFRFKKFCSSSTHNAECECIEG-FHCLGPQCTRCEKDCRPGQ-----ELTKQGCKTCSL
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
    TGHYSVRDQDRVYAG-----VSSTSSDFRDRPDGVSASEGPEGDPAGKEAEPAGPVSSLLGSPACVPIRA
      200          210          220          230          240          250

      130          140          150          160          170
    GTFNDQNGTGVCRPWTNCSLDGRSVLKTGTTEKDVVCGP-PV-----VSFSP-ST-TISVT----PEG
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
    GLGWVRDG-PRPHPYHFPAGSGGSLLRSAST---PVQGPVPVDLAPRQEEEEENQSPDSTEEEPVTVPRHTSD
    260          270          280          290          300          310          320

      180          190          200          210          220          230          240
    GPGGHSLLQVLTFLALTSALLLALIFITLLFSVLKWIRKKFPHIFKQPFKKTGAQEEEDACSCRCPEEEEG
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
    ADGFHLLKAGQSCFALIS--GSANQVKCYRFRVKKNHRHRYENCTTTSF---TVA----DNGAERQGGQAIL
      330          340          350          360          370          380

    250      X
      GGGGYEL
      :
      ITFGSPGQRQDFLKHVP
    390      X 400
  
```

## 10. ELLIS-267-3A

JL0104 Lymphocyte-associated cell surface molecule - Huma

ENTRY JL0104 #Type Protein  
 TITLE Lymphocyte-associated cell surface molecule - Human  
 SOURCE Homo sapiens #Common-name man  
 ACCESSION JL0104  
 REFERENCE



#Authors Tedder R. F., Isaacs C. M., Ernst T. J., Delmetti G. D.,  
 Adler D. A., Disteche C. M.  
 #Journal J. Exp. Med. (1989) 170:123-133  
 #Title Isolation and chromosomal localization of cDNAs  
 encoding a novel human lymphocyte cell surface  
 molecule, LAM-1. Homology with the mouse  
 lymphocyte homing receptor and other human  
 adhesion proteins.  
 #Molecule-type mRNA  
 #Residues 1-385 (TED)  
 #Comment The sequence shown here is composed of multi  
 homologous domains. One domain is homologous with  
 animal lectins, one is homologous with epidermal  
 growth factor, and two short consensus repeat  
 units similar to those found in C3/C4 binding  
 proteins.

# GENETIC

#Map-position 1q22-25  
 KEYWORDS membrane protein\ glycoprotein\ adhesion protein  
 FEATURE

1-51 #Domain signal sequence (predicted)  
 (SIG)\  
 52-385 #Protein lymphocyte-associated cell  
 surface molecule (predicted) (MAT)\  
 52-345 #Domain extracellular (probable) (EXT)\  
 346-368 #Domain transmembrane (probable) (TMM)\  
 369-385 #Domain cytoplasmic tail (CYT)\  
 73,117,190,245,259,  
 284,324 #Binding-site carbohydrate (Asn)  
 (potential)\  
 377,380 #Modified-site phosphorylation (Ser)  
 (probable)

SUMMARY #Molecular-weight 43743 #Length 385 #Checksum 4445  
 SEQUENCE

Initial Score = 9 Optimized Score = 49 Significance = 4.24  
 Residue Identity = 22% Matches = 66 Mismatches = 174  
 Gaps = 50 Conservative Substitutions = 0

```

      X      10      20      30      40      50
      MGNNCYNVVVIVLLLLVGCEKVG---AVQNSCDNCQPGTFCRKYNPVCKSCPPSTFSSIGGQP
      |      |      ||      |      |      |
AEIEYLEKTLPFSSYYWIGIRKIGGIWTWVGNTKSLTEEAENWGDGEPNNKKNKEDCVEIYIKRNKDAGKW
      90  X  100      110      120      130      140      150

60      70      80      90      100      110
N---CNICRVCAGYFRFKKFCSSSTHNAEC-ECIEGFHC-----LGPQC---TRCEKDCRPQGELTKQGCKT
|      |      |      |      |      |      |      |      |      |      |
NDDACHKLKAALCYTASCQPWSCSGHGECVEIINNYTCNCDVGYGYPQCQFVIQCEPLEAP--ELGTMDC-T
      160      170      180      190      200      210      220

120      130      140      150      160      170      180
CSLGTFN-DQNGTGVCRPWTNCSLDGRSVLKTGTTEKDVVCGPPVVSFSPSTTISV-----TPEGGPGGH
||      ||      ||      ||      ||      ||      ||
HPLGNFNFNSQCAFSCSEGTN--LTG----IEETT-----CEPFGNWSSPEPTCQVIQCEPLSAPDLGIMNC
      230      240      250      260      270      280

      190      200      210      220      230      240
SLQVLTLFLALTSALLLALIFITLLFSVLKWI-----RKKFPHIFKQPFKKTGAAQEEEDACSCRCP-----
|      |      |      |      |      |      |      |      |      |      |
S-HPLASF-SFTSACTFICSEGTELIGKKKTICSSGIWSNPSPICQKLDKSFSMIKEGDYNPLFIPVAVMV
      290      300      310      320      330      340      350

```

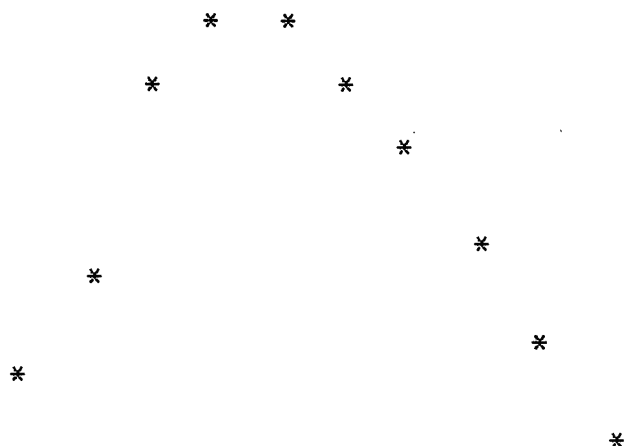
250 X  
 QEEEGGGGGYEL  
 | |

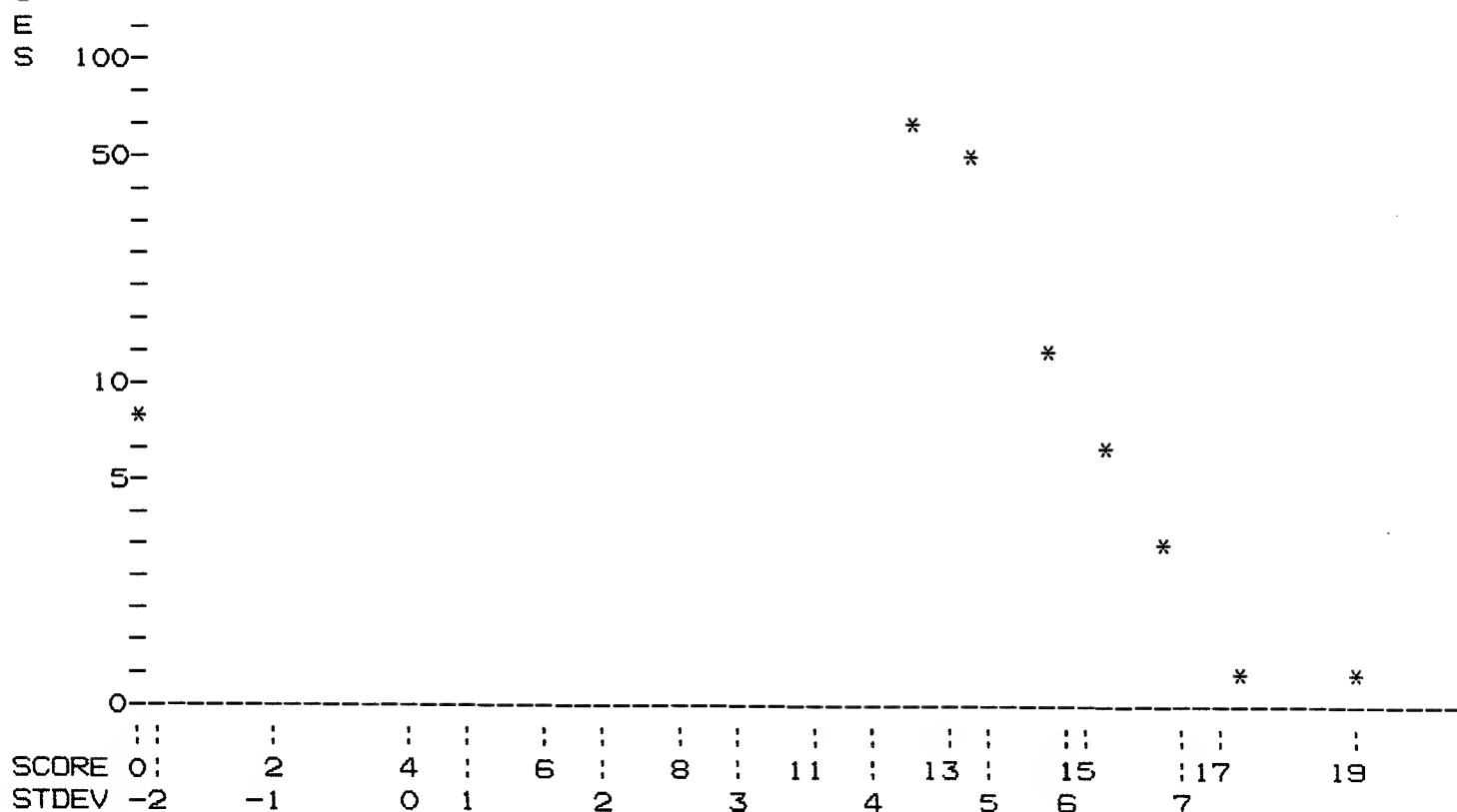
Results file ellis-267-3a-spt:res made by wendyc on Mon 27 Aug 90 16:08:31-PDT.

Query sequence being compared: ELLIS-267-3A  
 Number of sequences searched: 15409  
 Number of scores above cutoff: 4274

Results of the initial comparison of ELLIS-267-3A with:  
 Data bank : Swiss-Prot 14, all entries

10000-  
 -  
 N -  
 U 5000-  
 M -  
 B -  
 E -  
 R -  
 -  
 O -  
 F 1000-  
 -  
 S -  
 E 500-  
 Q -  
 U -  
 E -  
 N -





#### PARAMETERS

Similarity matrix	Unitary	K-tuple	2
Mismatch penalty	1	Joining penalty	20
Gap penalty	1.00	Window size	32
Gap size penalty	0.05		
Cutoff score	5		
Randomization group	0		
Initial scores to save	20	Alignments to save	10
Optimized scores to save	20	Display context	10

#### SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	5	7	1.78

Times:	CPU	Total Elapsed
	00:02:39.98	00:08:07.00

Number of residues:	4914263
Number of sequences searched:	15409
Number of scores above cutoff:	4274

Cut-off raised to 6.  
 Cut-off raised to 7.  
 Cut-off raised to 8.

The scores below are sorted by initial score.  
 Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Init. Length	Opt. Score	Sig. Frame
---------------	-------------	--------------	------------	------------

Best Available Copy  
 \*\*\*\* 7 standard deviations above mean \*\*\*\*

1.	HMO2\$HUMAN	OCTAMER BINDING TRANSCRIPTION	478	19	31	7.86	0
**** 6 standard deviations above mean ****							
2.	C1S\$HUMAN	COMPLEMENT COMPONENT C1S PRECU	688	17	43	6.73	0
3.	ECHM\$RAT	ENOYL-COA HYDRATASE, MITOCHOND	290	16	39	6.17	0
4.	MTDM\$MOUSE	DNA (CYTOSINE-5)-METHYLTRANSFE	1573	16	26	6.17	0
5.	TENA\$CHICK	TENASCIN (FRAGMENT).	697	16	41	6.17	0
**** 5 standard deviations above mean ****							
6.	COA1\$BFDV	COAT PROTEIN VP1.	343	15	30	5.61	0
7.	CADP\$MOUSE	PLACENTAL-CADHERIN PRECURSOR (	822	15	42	5.61	0
8.	KC2A\$DROME	CASEIN KINASE II, ALPHA CHAIN	335	15	26	5.61	0
9.	NIFA\$AZOVI	NIF-SPECIFIC REGULATORY PROTEI	522	15	38	5.61	0
10.	ATPX\$ANASP	ATP SYNTHASE B' CHAIN (EC 3.6.	163	15	22	5.61	0
11.	VE6\$HPV16	E6 PROTEIN.	158	15	28	5.61	0
12.	DHAS\$ECOLI	ASPARTATE-SEMIALDEHYDE DEHYDRO	367	14	25	5.05	0
13.	COX1\$SCHPO	CYTOCHROME C OXIDASE POLYPEPTI	537	14	43	5.05	0
14.	ENV\$BLV	ENV POLYPROTEIN (CONTAINS: COA	515	14	44	5.05	0
15.	LYAG\$HUMAN	LYSOSOMAL ALPHA-GLUCOSIDASE PR	951	14	41	5.05	0
16.	AST3\$DROME	ACHAETE-SCUTE COMPLEX PROTEIN	257	14	37	5.05	0
17.	SKI\$HUMAN	SKI ONCOGENE (GENE NAME: SKI).	728	14	40	5.05	0
18.	PRTZ\$BOVIN	PROTEIN Z.	396	14	34	5.05	0
19.	RRPL\$VSVSJ	RNA POLYMERASE BETA SUBUNIT (E	2109	14	43	5.05	0
20.	MYSG\$CHICK	MYOSIN HEAVY CHAIN, GIZZARD SM	1978	14	35	5.05	0

The scores below are sorted by optimized score.  
 Significance is calculated based on optimized score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
**** 5 standard deviations above mean ****						
1.	EGF\$MOUSE	EPIDERMAL GROWTH FACTOR PRECUR	1217	9	51	5.04 0
2.	VGLM\$PTPV	M POLYPROTEIN PRECURSOR (CONTA	1313	9	51	5.04 0
**** 4 standard deviations above mean ****						
3.	GMP1\$HUMAN	GRANULE MEMBRANE PROTEIN 140 P	830	7	50	4.58 0
4.	ALC\$RABIT	IG ALPHA CHAIN C REGION (FRAGM	299	9	50	4.58 0
5.	KAL\$HUMAN	PLASMA KALLIKREIN PRECURSOR (E	638	8	50	4.58 0
6.	OX40\$RAT	OX40 ANTIGEN PRECURSOR.	271	12	50	4.58 0
7.	CA36\$CHICK	COLLAGEN ALPHA 3(VI) (GENE NAM	2914	7	49	4.12 0
8.	LDLR\$HUMAN	LOW-DENSITY LIPOPROTEIN (LDL)	860	7	49	4.12 0
9.	RINI\$PIG	RIBONUCLEASE INHIBITOR.	456	9	49	4.12 0
10.	LAM1\$HUMAN	LEUKOCYTE ADHESION MOLECULE-1	372	9	49	4.12 0
11.	NGFR\$HUMAN	NERVE GROWTH FACTOR RECEPTOR P	427	8	49	4.12 0
12.	ATPB\$IPOBA	ATP SYNTHASE BETA CHAIN (EC 3.	489	10	49	4.12 0
13.	SUWA\$DROME	SUPPRESSOR-OF-WHITE-APRICOT PR	964	7	49	4.12 0
14.	ACDS\$HUMAN	ACYL-COA DEHYDROGENASE PRECURS	412	13	49	4.12 0
15.	CAML\$MOUSE	NEURAL CELL ADHESION MOLECULE	1260	7	49	4.12 0
16.	POLS\$SINDV	STRUCTURAL POLYPROTEIN (CONTAI	1245	13	49	4.12 0
17.	LNHR\$HUMAN	LYMPH NODE HOMING RECEPTOR PRE	372	7	49	4.12 0
18.	NEU\$RAT	NEU ONCOGENE PRECURSOR (EC 2.7	1260	9	49	4.12 0
19.	CHIT\$PHAVU	ENDOCHITINASE PRECURSOR (EC 3.	328	7	49	4.12 0
**** 3 standard deviations above mean ****						
20.	HEMA\$SENDH	HEMAGGLUTININ-NEURAMINIDASE (E	576	7	48	3.66 0

1. ELLIS-267-3A  
 EGF\$MOUSE EPIDERMAL GROWTH FACTOR PRECURSOR (EGF).

ID EGF\$MOUSE STANDARD; PRT; 1217 AA.  
 AC P01132;



```

ICRVC-----AGYFRFKFC---SSTHNAECEIEGFHCLGPQCTRCEKDCR-----PGQELTKQGCKTC
:      :      :      :      :      :      :      :      :      :      :      :
SYPGCPSSYDGYCLNGGVCMHIESLDSYTCNCVIGYSGDRCG-TR---DLRWELRHAGYGGQKHDIMVVAVC
980      990      1000      1010      1020      1030      1040

      120      130      140      150      160      170
----SLGTFNDQNGTGVC-----PWTN-CSLDGRSVLKTGTTEKDVVCGPPVVSFSPSTTISVTPEGGP
:      :      :      :      :      :      :      :      :      :      :
MVALVLLLLLGMWGTYYYYRTRKQLSNPPKNPCDEPSGSVSSSG---PDSSSGAAVASCPQWPFVLEKHQDP
1050      1060      1070      1080      1090      1100      1110

180      190      200      210      220      230      240
GGHSLQVLTLFLALTSALLLALIFITLLFSV-LKWIRKKFPHIFKQPFKKTGAAQEEEDACSCRCPQEEEG-
:      :      :      :      :      :      :      :      :      :      :
KNGSLPADGTNGAVVDA---GLSPSLQLGVSVHLTSWRQK-PHI---DGMGTGQSCWIPPSSDRGPQIEGN
1120      1130      1140      1150      1160      1170

      250      X
-----GGGGYEL
:      :
SHLPSYRPVGPEKLHSLQSANGS
1180      1190      1200

```

## 2. ELLIS-267-3A

VGLM\$PTPV M POLYPROTEIN PRECURSOR (CONTAINS: NONSTRUCTURAL P

```

ID  VGLM$PTPV      STANDARD;      PRT;  1313 AA.
AC  P03517;
DT  21-JUL-1986   (REL. 01, CREATED)
DT  21-JUL-1986   (REL. 01, LAST SEQUENCE UPDATE)
DT  01-OCT-1989   (REL. 12, LAST ANNOTATION UPDATE)
DE  M POLYPROTEIN PRECURSOR (CONTAINS: NONSTRUCTURAL PROTEIN NS-M;
DE  GLYCOPROTEINS G1 AND G2).
OS  PUNTA TORO PHLEBOVIRUS.
OC  VIRIDAE; SS-RNA ENVELOPED VIRUSES; BUNYAVIRIDAE.
RN  [1] (SEQUENCE FROM N. A.)
RA  IHARA T., SMITH J., DALRYMPLE J.M., BISHOP D.H.L.;
RL  VIROLOGY 144:246-259(1985).
CC  -!- SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS
CC  INCLUDING NONSTRUCTURAL PROTEIN NS-M, GLYCOPROTEIN G1, AND
CC  GLYCOPROTEIN G2.
DR  EMBL; M11156; PTPMRNA.
DR  PIR; A04109; VGVUPT.
KW  POLYPROTEIN; GLYCOPROTEIN; TRANSMEMBRANE; NONSTRUCTURAL PROTEIN.
FT  CHAIN          1      270      NONSTRUCTURAL PROTEIN NS-M.
FT  CHAIN          271      809      GLYCOPROTEIN G1.
FT  CHAIN          810     1313      GLYCOPROTEIN G2.
FT  CARBOHYD        76       76      POTENTIAL.
FT  CARBOHYD       102      102      POTENTIAL.
FT  CARBOHYD       496      496      POTENTIAL.
FT  CARBOHYD      1154     1154      POTENTIAL.
FT  CARBOHYD      1243     1243      POTENTIAL.
SQ  SEQUENCE       1313 AA;  146374 MW;  9199811 CN;

```

```

Initial Score      =      9   Optimized Score    =     51   Significance =   5.04
Residue Identity   =     23%   Matches            =     68   Mismatches  =   172
Gaps               =     53   Conservative Substitutions =      0

```

```

      X      10      20      30      40      50
      MGNNCYNVVVIVLLLLVGCEKVGAVQNSCDNCQPG--TFCKRY-----NPVCKSCPPSTFS
      :      :      :      :      :      :      :      :      :      :
TNVSFVCYEHVGGDEQEVHRALKRVSVDCKIVDNSKQKICTGDHVFCEKYDCSTSYDPDVTCTIHAPGSGPL
500      X  510      520      530      540      550      560

```

Best Available Copy

```

S I G G G P N C N I C R V C A G Y ----- F R F K K F C S S T H N A E C ----- E C - I E G F H C L G P Q C T R C E K D C R P G Q E L T - K
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
Y I - N L M G S W I K P Q C V G Y E R V L V D R E V K Q P L L A P E Q N C D T C V S E C L D E G V H ----- I K S T G F E I T S A
570           580           590           600           610           620

      120      130      140      150      160      170      180
Q G C K T C S L G T F N D Q N G T -- G V C R P W T N C S L D G R S V L K T G T T E K D V V C G P P V V S F S P -- S T T I S V T P E G G P G G
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
V A C S H G S C I S A H Q E P S T S V I V P Y P G L L A S V G G R I G I H L S H T - S D S A S V H M V V V C P P R D S C A A H N C L L C Y H G I
630           640           650           660           670           680           690

      190      200      210      220      230      240
H S L Q - V L T L F L A L T S A L L L A L I F I T L L F S V L K W I R K K F P H I F K Q P F K K T T G ----- A A Q E E D A C S C
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
L N Y Q C H S T L S A I L T S F L L -- I L F I Y T V F S V T T N I L Y V L R L I P K Q - L K S P V G W L K L F I N W L L T A L R I K T R N V M
700           710           720           730           740           750           760

      250      X
R C P Q E E E G G G G G Y E L
|   |
R R I N Q R I G W V D H H D V E R P R H R E P M R
770           780           790

```

### 3. ELLIS-267-3A

GMP1\$HUMAN GRANULE MEMBRANE PROTEIN 140 PRECURSOR.

```

ID  GMP1$HUMAN      STANDARD;      PRT;      830 AA.
AC  P16109;
DT  01-APR-1990 (REL. 14, CREATED)
DT  01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
DT  01-APR-1990 (REL. 14, LAST ANNOTATION UPDATE)
DE  GRANULE MEMBRANE PROTEIN 140 PRECURSOR.
OS  HUMAN (HOMO SAPIENS).
OC  EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC  EUTHERIA; PRIMATES.
RN  [1] (SEQUENCE FROM N. A.)
RA  JOHNSTON G. I. , COOK R. G. , MCEVER R. P. ;
RL  CELL 56:1033-1044(1989).
DR  PIR; A30359; A30359.
DR  EMBL; M25322; M25322.
DR  PROSITE; PS00022; EGF.
KW  LECTIN; GLYCOPROTEIN; TRANSMEMBRANE; SIGNAL.
FT  SIGNAL          1          41
FT  CHAIN           42         830      GRANULE MEMBRANE PROTEIN 140.
FT  DOMAIN          42         159      LECTIN.
FT  DOMAIN          160        199      EGF-LIKE.
FT  DOMAIN          200        770      COMPLEMENT H/C4B-BINDING.
FT  TRANSMEM        772        795      PUTATIVE.
FT  CARBOHYD         54         54      PUTATIVE.
FT  CARBOHYD         98         98      PUTATIVE.
FT  CARBOHYD        180        180      PUTATIVE.
FT  CARBOHYD        212        212      PUTATIVE.
FT  CARBOHYD        219        219      PUTATIVE.
FT  CARBOHYD        411        411      PUTATIVE.
FT  CARBOHYD        460        460      PUTATIVE.
FT  CARBOHYD        518        518      PUTATIVE.
FT  CARBOHYD        665        665      PUTATIVE.
FT  CARBOHYD        716        716      PUTATIVE.
FT  CARBOHYD        723        723      PUTATIVE.
FT  CARBOHYD        741        741      PUTATIVE.
SQ  SEQUENCE        830 AA;  90766 MW;  3510536 CN;

```

```

Initial Score      =      7      Optimized Score      =      50      Significance      =      4.58
Residue Identity   =     22%      Matches              =      66      Mismatches       =     174

```

gaps = 47 Conservative Substitutions = 0

Best Available Copy

```

      X      10      20      30      40      50      60
      MGNNCYNVVVIVLLLLVGCEKVGAVQNSCDNCQPGTFCRKYNPVCKSCPPSTFSSIGGQPNCN
      :      :      :      :      :      :      :      :      :      :
NEARVNC SHPFGAFRYQSVCSFTCNEGLLLL VGA---SVLQCLATGNWNSVPPECQAIPCTPLLS---PQNGTM
      460      X      470      480      490      500      510      520

      70      80      90      100      110      120
      ICRVCAGYFRFKKFCSSSTHNAECECIEGFHCLGP---QCTR---CEKDCRPGQELTK-----QGCKTCS-
      :      :      :      :      :      :      :      :      :      :
      TCVQPLGSSSYKSIC-----QFICDEGYSLSGPERLDCTRSGRWTDSPPMCEAIKCPELFAPEQGSLLDCSD
      530      540      550      560      570      580

      130      140      150      160      170      180
      -LGTFN-----DQNGTGVCRP-WTNCSLDGRSVLKTGTTEKDVVCGPPVVSFSPSTTISVTPEGGPGGH
      :      :      :      :      :      :      :      :      :      :
      TRGEFNVGSTCHFSCNNGFKLEGPNNVECTTSGR-WSATPPTCKGIASLPTPGLQCPALT---TPGQGTMYC
      590      600      610      620      630      640      650

      190      200      210      220      230      240
      SLQVLTLFLALTSALLLALIFITLLFSVLKWIRKKFPHIFKQP---FKKTTGAQGEEDA---CSCRCPEEE
      :      :      :      :      :      :      :      :      :      :
      RHHPGT-FGFNTTCYFGCNAGFTLIGDSTLSCRPSGQWTAVTPACRAVKCSELHVNKPIAMNCSNLWGNFSY
      660      670      680      690      700      710      720

      250      X
      G-GGGGYEL
      :      :      :
      GSICSGHCLEGQLLNGSAQ
      730      X      740

```

#### 4. ELLIS-267-3A

ALC\$RABIT IG ALPHA CHAIN C REGION (FRAGMENT).

```

ID  ALC$RABIT      STANDARD;      PRT;      299 AA.
AC  P01879;
DT  21-JUL-1986 (REL. 01, CREATED)
DT  21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
DT  01-NOV-1988 (REL. 09, LAST ANNOTATION UPDATE)
DE  IG ALPHA CHAIN C REGION (FRAGMENT).
OS  RABBIT (ORYCTOLAGUS CUNICULUS).
OC  EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC  EUTHERIA; LAGOMORPHA.
RN  [1] (SEQUENCE FROM N. A.)
RA  KNIGHT K. L. , MARTENS C. L. , STOKLOSA C. M. , SCHNEIDERMAN R. D. ;
RL  NUCLEIC ACIDS RES. 12:1657-1670(1984).
CC  -!- THIS IMMUNOGLOBULIN BELONGS TO THE IGA-G SUBCLASS. IT WAS ISOLATED
CC  FROM A RABBIT HOMOZYGOUS FOR A2, N80, DE12,15, F71, G75 HEAVY
CC  CHAIN HAPLOTYPE.
DR  PIR; A02174; AHRB.
DR  EMBL; X00353; OCIG02.
DR  PROSITE; PS00290; IG_MHC.
KW  IMMUNOGLOBULIN C REGION.
FT  NON_TER      1      1
SQ  SEQUENCE      299 AA; 32256 MW; 500462 CN;

```

Initial Score = 9 Optimized Score = 50 Significance = 4.58  
 Residue Identity = 23% Matches = 69 Mismatches = 153  
 Gaps = 70 Conservative Substitutions = 0

```

      X      10      20      30      40      50      60
      MGNNCYNVVVIVLLLLVGCEKVGAVQNSCDNCQPGTFCRKYNPVCKSCPPSTFSSIGGQPNC---NICRVCA
      :      :      :      :      :      :      :      :      :      :
      QSGTSGPYTACSELILPVTQCLG---QKS-AAC-----HVEYNSVINESLPVPF-----PDCCPANSCCTC-

```



Best Available Copy

```

70      80      90      100      110      120      130
GYFRFKKFCSSSTHNAECECIEGFHCLGPQCTRCEKDCRPGQELTKQCK--TCSLGTFNQNGTGVCRPWTN
      : : : : : : : : : : : : : : : : : :
-----PSSSSRNLI SGCQPSLSLQRPDLGDL LGRDASLTCTLSGLKNPEDAVFTWEPTNGNEPVQQRAQ
      60      70      80      90      100      110      120

140      150      160      170      180      190
CSLDG---RSVL-----KTGTTEKDVVCGPPVVSFSPSTTIS---VTPEGGPGGHSLQVLTLFLAL TSA
      : : : : : : : : : : : : : : : : : :
RDLSGCYSVSSVLPSSAETWKARTEFTCTVTHPEIDSGSLTATISRGVVTP---PQVHLLPPPSEELALNEQ
      130      140      150      160      170      180      190

200      210      220      230      240      250
LLLALIFITLL---FS---VLKWIR-----KKFPHIFKQP---FKKTTGAQDEEDACSCRCPEEEEGGG
      : : : : : : : : : : : : : : : : : :
VTL-----TCLVRGFSPKDVLSWRHQGGQEVPEDSFLVWKSMPESQDKATYA----ITSLLRVPAEDWNQ
      200      210      220      230      240      250

      X
GGYEL
      :
DTYSCMVGHEGLAEH
      260

```

5. ELLIS-267-3A

KAL\$HUMAN PLASMA KALLIKREIN PRECURSOR (EC 3. 4. 21. 34) (PLASMA

```

ID  KAL$HUMAN      STANDARD;      PRT;      638 AA.
AC  P03952;
DT  23-OCT-1986 (REL. 02, CREATED)
DT  23-OCT-1986 (REL. 02, LAST SEQUENCE UPDATE)
DT  01-APR-1990 (REL. 14, LAST ANNOTATION UPDATE)
DE  PLASMA KALLIKREIN PRECURSOR (EC 3. 4. 21. 34) (PLASMA PREKALLIKREIN)
DE  (KININOGENIN).
OS  HUMAN (HOMO SAPIENS).
OC  EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC  EUTHERIA; PRIMATES.
RN  [1] (SEQUENCE FROM N. A.)
RA  CHUNG D. W. , FUJIKAWA K. , MCMULLEN B. A. , DAVIE E. W. ;
RL  BIOCHEMISTRY 25:2410-2417(1986).
CC  -!- FUNCTION: THE ENZYME CLEAVES LYS-ARG AND ARG-SER BONDS. IT
CC      ACTIVATES, IN A RECIPROCAL REACTION, FACTOR XII AFTER ITS BINDING
CC      TO A NEGATIVELY CHARGED SURFACE. IT ALSO RELEASES BRADYKININ FROM
CC      HMW KININOGEN AND MAY ALSO PLAY A ROLE IN THE RENIN-ANGIOTENSIN
CC      SYSTEM BY CONVERTING PRORENIN INTO RENIN.
CC  -!- SUBUNIT: THE ZYMOGEN IS ACTIVATED BY FACTOR XIIA, WHICH CLEAVES
CC      THE MOLECULE INTO A LIGHT CHAIN, WHICH CONTAINS THE ACTIVE SITE,
CC      AND A HEAVY CHAIN, WHICH ASSOCIATES WITH HMW KININOGEN. THESE
CC      CHAINS ARE LINKED BY ONE OR MORE DISULFIDE BONDS.
DR  EMBL; M13143; HSPPKKA.
DR  PIR; A00921; KQHUP.
DR  PROSITE; PS00134; TRYPSIN_HIS.
DR  PROSITE; PS00135; TRYPSIN_SER.
KW  HYDROLASE; SERINE PROTEASE; GLYCOPROTEIN; PLASMA; ZYMOGEN; SIGNAL;
KW  FIBRINOLYSIS; BLOOD COAGULATION; INFLAMMATION; LIVER; DUPLICATION;
KW  BRADYKININ.
FT  SIGNAL          1          19
FT  CHAIN           20         390      PLASMA KALLIKREIN, HEAVY CHAIN.
FT  CHAIN          391         638      PLASMA KALLIKREIN, LIGHT CHAIN.
FT  DOMAIN         389         621      SERINE PROTEASE.
FT  REPEAT          20         104
FT  REPEAT         110         194
FT  REPEAT         200         284

```

Best Available Copy

FT	CARBOHYD	127	127	
FT	CARBOHYD	308	308	
FT	CARBOHYD	396	396	
FT	CARBOHYD	453	453	
FT	CARBOHYD	494	494	
FT	ACT_SITE	434	434	CHARGE RELAY SYSTEM.
FT	ACT_SITE	483	483	CHARGE RELAY SYSTEM.
FT	ACT_SITE	578	578	CHARGE RELAY SYSTEM.
SQ	SEQUENCE	638 AA;	71369 MW;	2175970 CN;

Initial Score = 8 Optimized Score = 50 Significance = 4.58  
 Residue Identity = 23% Matches = 70 Mismatches = 164  
 Gaps = 69 Conservative Substitutions = 0

```

      X              10              20              30              40              50
      MGNNC-----YNVVVI-----VLLL VGCEKVGAVQNSCDNCQPGTFCRKYNPVCKSCPST
      |||          |||          |||          |||          |||          |||
DAFVCRITICTYHPNCLFFTFYTNVWKIESQRNVCLLKTSE-SGTPSSS--TPQENTISGYSLLTCKRTLPEP
      230 X      240      250      260      270      280      290

      60              70              80              90              100             110
      FSS---IGGQPN CNICRV CAGYFRFK---FCSSTHNAECECIEGFHCLGPQCTRCEKDCRPGQELTKQGCK
      |||          |||          |||          |||          |||          |||
CHSKIYPGVDFGGEELNV-----TFVKGVNVQDETCTKMIRCQFFTYSLLPEDCKEEK-CKCFLRLSMDGSP
      300      310      320      330      340      350

      120             130             140             150             160             170
      T-----CSLGTFNQNGTG---VCRPWTNCSLDGRSVLKTGTTEKDVVCGPPVVSFSPSTTISVTPEGGP
      |||          |||          |||          |||          |||          |||
TRIAYGTQSSGYSLRLCNTGDNVCTTKT---STRIV---GGTNSSWGEWPWQVSLQVKLTAQRHLCGGS
      360      370      380      390      400      410      420

      180             190             200             210             220             230
      GGHSLLQVLT-----LFL---ALTSALL-LALIFITLLFSVLKWIRKKFPHIFKQPFKKTG---AAGE
      |||          |||          |||          |||          |||          |||
LIGHQWVLTAAHCFDGLPLQDVWRIYSGILNLSDITKDTPFSQIKEI-----IIHQNYKVSEGNHDIALIK
      430      440      450      460      470      480

      240             250             X
      EDA---CSCRCPQEEEGGGGGYEL
      |||          |||          |||
LQAPLNYTEFQKPICLPSKGDSTIYTNCWVTGWG
      490      500      510 X      520

```

# 6. ELLIS-267-3A

OX40\$RAT OX40 ANTIGEN PRECURSOR.

ID OX40\$RAT PRELIMINARY; PRT; 271 AA.  
 AC P15725;  
 DT 01-APR-1990 (REL. 14, CREATED)  
 DT 01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)  
 DT 01-APR-1990 (REL. 14, LAST ANNOTATION UPDATE)  
 DE OX40 ANTIGEN PRECURSOR.  
 OS RAT (RATTUS NORVEGICUS).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; RODENTIA.  
 RN [1] (T-CELL, SEQUENCE FROM N. A.)  
 RA MALLETT S., FOSSUM S., BARCLAY A.N.;  
 RL SUBMITTED (OCT-1989) TO EMBL/GENBANK DATA BANKS.  
 CC -!- SIMILARITY: TO NERVE GROWTH FACTOR RECEPTOR.  
 DR EMBL; X17037; RSOX40.  
 KW T-CELL; ANTIGEN; GLYCOPROTEIN; SIGNAL.  
 FT SIGNAL 1 19  
 FT CHAIN 20 271 OX-40 ANTIGEN.

FT REPEAT 25 60 Cysteine-rich repeat I.  
 FT REPEAT 61 102 Cysteine-rich repeat II.  
 FT REPEAT 123 164 Cysteine-rich repeat III.  
 SQ SEQUENCE 271 AA; 29895 MW; 400796 CN;

Initial Score = 12    Optimized Score = 50    Significance = 4.58  
 Residue Identity = 25%    Matches = 72    Mismatches = 145  
 Gaps = 63    Conservative Substitutions = 0

```

      X      10      20      30      40      50
      MGNNCYNVVVIVLLLVGCE-KVGAVQNS-CD-----NCQPGTFCRKYN-PVCKSCPST
      ||      |      |      |      |      |      |      |      |
    LLLGLSLGVTVKLNCVKDTPSGHKCCRECGPHGMVSRCDHTRDTVCHPCEPGFYNEAVNYDTCKQC----
      20 X      30      40      50      60      70

      60      70      80      90      100      110
    FSSIGGQPNICRVCAGYFRFKKFCSSTHNAECECIEG-----FHCLGPQCTRC-EKDCRPGQELTKQG
      |      |      |      |      |      |      |      |      |      |
    -----TQCN---HRS---GSELKQNCPTEDTVCGCRPGTGPRQDSSHLGVDCVPCPPGHFSPG---SNQA
      80      90      100      110      120      130

      120      130      140      150      160      170
    CKTCSLGTFNQNGTGVCRPWTNCSLD-----GRSVLKT-----GTT--EKDV--VCGPPVVSFSPSTTISV
      |      |      |      |      |      |      |      |      |      |
    CK---PWTNCTLSGKQIRHPASN-SLDTVCEDRSLLATLLWETQRTTFRPTTVPSTTVWPRTSQLPSTPTLV
      140      150      160      170      180      190      200

      180      190      200      210      220      230      240
    TPEGGPGGHSLSQVLTFLALTSALLLALIFITLLFSVLKWIRKKFPHIFKQPFKKTGAQGEEDACSCRCPQ
      ||      |      |      |      |      |      |      |      |      |
    APE-GPAFAVILGLGLLAPLTVLLAL---YLL--RKAWRSPNTPKPCWGNSFRT--PIQEEQTDTHFTLA
      210      220      230      240      250      260
  
```

X  
 EEEGGGGGYEL

KI  
 270

#### 7. ELLIS-267-3A

CA36#CHICK COLLAGEN ALPHA 3(VI) (GENE NAME: COL6A3) (FRAGMENT

ID CA36#CHICK STANDARD; PRT; 2914 AA.  
 AC P15989;  
 DT 01-APR-1990 (REL. 14, CREATED)  
 DT 01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)  
 DT 01-APR-1990 (REL. 14, LAST ANNOTATION UPDATE)  
 DE COLLAGEN ALPHA 3(VI) (GENE NAME: COL6A3) (FRAGMENT).  
 OS CHICKEN (GALLUS GALLUS).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; AVES.  
 RN [1] (SEQUENCE FROM N. A.)  
 RA BONALDO P., RUSSO V., BUCCIOTTI F., DOLIANA R., COLOMBATTI A.;  
 RL SUBMITTED (SEP-1989) TO EMBL/GENBANK DATA BANKS.  
 RN [2] (SEQUENCE OF 2648-2914 FROM N. A.)  
 RA BONALDO P., COLOMBATTI A.;  
 RL J. BIOL. CHEM. 264:20235-20239(1989).  
 CC -!- FUNCTION: COLLAGEN VI ACTS AS A CELL-BINDING PROTEIN.  
 CC -!- SUBUNIT: TRIMERS COMPOSED OF THREE DIFFERENT CHAINS: ALPHA 1(VI),  
 CC ALPHA 2(VI), AND ALPHA 3(VI).  
 CC -!- PROLINES IN THE THIRD POSITION OF THE TRIPEPTIDE REPEATING UNIT  
 CC (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 DR EMBL; M24282; GGCOLAVI.  
 KW EXTRACELLULAR MATRIX; CONNECTIVE TISSUE; TANDEM REPEAT; HYDROXYLATION;  
 KW GLYCOPROTEIN; CELL ADHESION.

SEQUENCE 2914 AA: 315788 MW: 2.213953E+07 CN;  
Best Available Copy

Initial Score = 7 Optimized Score = 49 Significance = 4.12  
Residue Identity = 23% Matches = 65 Mismatches = 168  
Gaps = 42 Conservative Substitutions = 0

```

      X          10          20          30          40          50
      MGN-NCYNV---VV-IV-LLL VGCEK--VGAVQNSCDNCQPGTFCRKYNPVCKSCPPSTFSS
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
II FL LDGSLNVGNANFPFVRDFVVTLVN YLDVGTDKIRVGLVQFS---DTPKTEFSLSYQTK----SDIIQ
      430   X   440          450          460          470          480

      60          70          80          90          100          110          120
      IGGQPN CNICRV CAGYFRFKKFCSS THNAE---CECIEGFHCLGPQCTRCEKDCRPGQ---ELTKQGGCKTCS
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      RL GQLRPKGGSV-LNTGSALNFVLSNH FTEAGGSRINEQVPQVLVLVTAGRS AVPFLQVSN DLARAGVLTFA
      490          500          510          520          530          540          550

      130          140          150          160          170          180
      LGTFN---DQNGTGVC RPWTNCSLDGRSVLKTGTTEKDVVCGPPVVSFSPSTTI-SVTPEGGP-GGHS LQVL
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      VGVRNADKAELEQIAFNPKMVYFMDDFSDLTT-----LPQELKKPITTIVSGGVEEVPLAPTESKKD
      560          570          580          590          600          610          620

      190          200          210          220          230          240          250          X
      TLFLALTSALLLALIFITLLFSVLKWIRKKFPHIFKQPFKKTGAAGQ EEDACSCRC PQEEEGGGGGYEL
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      ILFLIDGSANLL-----GSFPAVRDFIHKVISDLNVGP DATRVAVAQFSDNIQIEFDFAELPSKQDMLLKVK
      630          640          650          660          670          680          X

      RMRLKTG
      690
```

#### 8. ELLIS-267-3A

LDLR\$HUMAN LOW-DENSITY LIPOPROTEIN (LDL) RECEPTOR PRECURSOR.

ID LDLR\$HUMAN STANDARD; PRT: 860 AA.  
AC P01130;  
DT 21-JUL-1986 (REL. 01, CREATED)  
DT 21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)  
DT 13-AUG-1987 (REL. 05, LAST ANNOTATION UPDATE)  
DE LOW-DENSITY LIPOPROTEIN (LDL) RECEPTOR PRECURSOR.  
OS HUMAN (HOMO SAPIENS).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; PRIMATES.  
RN [1] (SEQUENCE FROM N. A.)  
RA YAMAMOTO T., DAVIS C. G., BROWN M. S., SCHNEIDER W. J., CASEY M. L.,  
RA GOLDSTEIN J. L., RUSSELL D. W.;  
RL CELL 39:27-38(1984).  
CC -!- THIS TRANSMEMBRANE GLYCOPROTEIN BINDS LDL, THE MAJOR CHOLESTEROL-  
CC CARRYING LIPOPROTEIN OF HUMAN PLASMA, & TRANSPORTS IT INTO CELLS  
CC BY ENDOCYTOSIS. IN ORDER TO BE INTERNALIZED, THE RECEPTOR-LIGAND  
CC COMPLEXES MUST FIRST CLUSTER INTO CLATHRIN-COATED PITS.  
CC -!- THE AMINO END OF THE EXTRACELLULAR DOMAIN CONTAINS 7 OR 8 40-  
CC RESIDUE REPEATS. EACH REPEAT HAS ABOUT 6 CYS RESIDUES, ALL OF  
CC WHICH ARE INVOLVED IN DISULFIDE BONDS. FOLLOWING THESE REPEATS IS  
CC A REGION OF ABOUT 350 RESIDUES THAT IS HOMOLOGOUS WITH PART OF THE  
CC EPIDERMAL GROWTH FACTOR (EGF) PRECURSOR.  
CC -!- THE LAST HALF OF THE EXTRACELLULAR DOMAIN CONTAINS STRUCTURAL  
CC EVIDENCE OF REPETITIVE SEQUENCE.  
CC -!- AN INTRAstrand RECOMBINATION EVENT BETWEEN TWO ALU SEQUENCES IN  
CC THE 3' UNTRANSLATED REGION OF MRNA FROM A FAMILIAL HYPERCHOLEST-  
CC EROLEMIA PATIENT RESULTS IN THE DELETION OF THE TRANSMEMBRANE &  
CC CYTOPLASMIC DOMAINS. MOST OF THE RECEPTORS PRODUCED ARE SECRETED,

CC BUT THOSE THAT ADHERE TO THE CELL SURFACE CANNOT CLUSTER IN COATED  
 CC PITS; THEREFORE, EVEN THOUGH THEY BIND LDL, THESE RECEPTOR-LIGAND  
 CC COMPLEXES ARE NOT INTERNALIZED.  
 DR PIR; A01383; QRHULD.  
 DR EMBL; K02573; HSLDLR.  
 KW GLYCOPROTEIN; LDL; CHOLESTEROL METABOLISM; LIPID TRANSPORT;  
 KW ENDOCYTOSIS; COATED PITS; TRANSMEMBRANE; RECEPTOR; SIGNAL.  
 FT SIGNAL 1 21  
 FT CHAIN 22 860 LDL RECEPTOR.  
 FT DOMAIN 22 788 EXTRACELLULAR.  
 FT TRANSMEM 789 810  
 FT DOMAIN 811 860 CYTOPLASMIC.  
 FT REPEAT 22 61 CYSTEIN RICH.  
 FT REPEAT 62 102 CYSTEIN RICH.  
 FT REPEAT 103 141 CYSTEIN RICH.  
 FT REPEAT 142 180 CYSTEIN RICH.  
 FT REPEAT 191 229 CYSTEIN RICH.  
 FT REPEAT 230 268 CYSTEIN RICH.  
 FT REPEAT 269 309 CYSTEIN RICH.  
 FT SIMILAR 311 661 WITH EGF PRECURSOR.  
 FT REPEAT 441 445  
 FT REPEAT 488 492  
 FT REPEAT 531 535  
 FT REPEAT 575 579  
 FT REPEAT 617 621  
 FT SITE 721 768 CLUSTERED O-LINKED OLIGOSACCHARIDES.  
 SQ SEQUENCE 860 AA; 95375 MW; 3807460 CN;

Initial Score = 7 Optimized Score = 49 Significance = 4.12  
 Residue Identity = 23% Matches = 67 Mismatches = 162  
 Gaps = 56 Conservative Substitutions = 0

```

X      10      20      30      40      50      60
MGNNCYNVVVIVLLLLVGCEKVGAVQNSC-----DNCQPGTFCRKYNPVCKSCPPSTFSSIGGQPNICRVCA
||      ||      ||      ||      ||      ||      ||      ||      ||      ||
MGPWGWLRLRTVALL-LAAAGTAVGDR CERNEFQCQDG-KCISYKWVCDGSAECQDGSDESQETCLSVTCKS
X      10      20      30      40      50      60      70
70      80      90      100     110     120     130
GYFRFKKFCSSSTHNAECECIEGFHCLGPQCTRC--EKDCRPGQELTKQGC--KTCSLGT FNDQNGTGVC RPW
||      ||      ||      ||      ||      ||      ||      ||      ||      ||
GDF----SCGGRVN---RCI-----PQFWRC DGQVDCDNG--SDEQGC PPKTCSQDEF RCHD GK CISRQF
      80      90      100     110     120
140      150     160     170     180     190
TNCS----LDG----RSVLKTGTTEKDVVCGPPVVSFSPSTTI-----SVTPEGGPGGHSLQVLTL-F
||      ||      ||      ||      ||      ||      ||      ||      ||      ||
VCDSDRDCLDGSDEASCPVL-----TCGPASFQCNSSTCIPQLWACDNDPDCEDGSDEWPQRCRGLYV
130      140      150     160     170     180
200      210     220     230     240     250     X
LALTSALLLALIFITL--LFSVLKWIRKKFP-HIFKQFFKKTGAAGQEDACSCRCPPQEEEGGGGGYEL
||      ||      ||      ||      ||      ||      ||      ||      ||      ||
FQGDSSPCSAFEFHCLSGECIHSSWRC DGGPDCKDKSDEENCAVATCRPDEFQCSDGNCIHGSRQCDREYDC
190      200     210     220     230     240     250     260
K DMSDEV

```

9. ELLIS-267-3A  
 RINI#PIG RIBONUCLEASE INHIBITOR.

ID RINI#PIG STANDARD; PRT; 456 AA.  
 AC P10775;  
 DT 01-JUL-1989 (REL. 11, CREATED)

DT 01-JUL-1989 (REL. 11, LAST SEQUENCE UPDATE)  
 DE RIBONUCLEASE INHIBITOR.  
 OS PIG (SUS SCROFA).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; ARTIODACTYLA.  
 RN [1] (LIVER, SEQUENCE)  
 RA HOFSTEENGE J. , KIEFFER B. , MATTHIES R. , HEMMINGS B. A. , STONE S. R. ;  
 RL BIOCHEMISTRY 27:8537-8544(1988).  
 CC -!- FUNCTION: THIS PROTEIN IS AN INHIBITOR OF PANCREATIC RNASE AND  
 CC ANGIOGENIN.  
 CC -!- THERE ARE 15 LEUCINE-RICH REPEATS.  
 CC -!- SIMILARITY: THE REPEATED LEUCINE-RICH SEGMENT IS FOUND IN MANY  
 CC PROTEINS.  
 KW ACETYLATION; TANDEM REPEAT; LEUCINE-REPEAT.  
 FT MOD\_RES 1 1 ACETYLATION.  
 FT REPEAT 15 43 A1.  
 FT REPEAT 44 71 B1.  
 FT REPEAT 72 100 A2.  
 FT REPEAT 101 128 B2.  
 FT REPEAT 129 157 A3.  
 FT REPEAT 158 185 B3.  
 FT REPEAT 186 214 A4.  
 FT REPEAT 215 242 B4.  
 FT REPEAT 243 271 A5.  
 FT REPEAT 272 299 B5.  
 FT REPEAT 300 328 A6.  
 FT REPEAT 329 356 B6.  
 FT REPEAT 357 385 A7.  
 FT REPEAT 386 413 B7.  
 FT REPEAT 414 442 A8.  
 SQ SEQUENCE 456 AA; 49022 MW; 991302 CN;

Initial Score = 9 Optimized Score = 49 Significance = 4.12  
 Residue Identity = 23% Matches = 68 Mismatches = 152  
 Gaps = 71 Conservative Substitutions = 0

```

      X      10      20      30      40      50
      MGNNCYNVVVIVLLLLVGCEKVGAVQNSCDNCQPGTFCRKYNPVCKSCP--PSTFS--SIGGQ
      ||      ||      ||      ||      ||      ||      ||      ||
  ADSACQLETLRLNCGLT PANCKDLCG---IVASQASLREL DLGSNGLGDAGIAELCPGLLSPASRLKTLWL
190      200      210      220      230      240      250

      60      70      80      90      100      110      120
      PNCNI----CRVCAGYFRFK---KFCSSSTHNAECECIEGFHCLGPQCTRCEKDCRPGQELTKQGCKTCSLGT
      ||      ||      ||      ||      ||      ||      ||      ||
  WECDITASGCRDLRCRVLQAKETLKELSLAGN--KLGDEGARLL-----CESLLQPGQLES LWKSCSLTA
260      270      280      290      300      310      320

      130      140      150      160      170      180
      FNDQNGTGVCRPWT-NCSL--DGRSVLKTGTTEKDVVCGPPVVSFSPSTTISVTPEG-----GPGGHSLQVL
      ||      ||      ||      ||      ||      ||      ||      ||
  ACCQ---HVSLMLTQNKHLLLELQSSNKLGDSGIQELC---QALSQPGTTLRVLC LGDCEVTNSGCSSL--A
      330      340      350      360      370      380

      190      200      210      220      230      240
      TLFLALTSALLLAL-----IFITLLFSVLKWIRKKFPHIFKQPFKKTGAAGQEDACSCRCPEE-----
      ||      ||      ||      ||      ||      ||      ||      ||
  SLLLANRSLRELDLSNNCVGDPGVLQLLGS-----LEQP-----GCALEQLVLYDTYWTEEEVEDR
      390      400      410      420      430      440

      250      X
      ----EGGGGGYEL
      ||      ||
  LQALEGSKPGLRVIS
      450      X
  
```

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ **BLACK BORDERS**
- ☒ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☒ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☒ **SKEWED/SLANTED IMAGES**
- ☒ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☒ **GRAY SCALE DOCUMENTS**
- ☒ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**

## 10. ELLIS-267-3A

LAM1\$HUMAN LEUKOCYTE ADHESION MOLECULE-1 PRECURSOR (LAM-1).

ID LAM1\$HUMAN PRELIMINARY; PRT; 372 AA.  
 AC P15023;  
 DT 01-APR-1990 (REL. 14, CREATED)  
 DT 01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)  
 DT 01-APR-1990 (REL. 14, LAST ANNOTATION UPDATE)  
 DE LEUKOCYTE ADHESION MOLECULE-1 PRECURSOR (LAM-1).  
 OS HUMAN (HOMO SAPIENS).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; PRIMATES.  
 RN [1] (TONSIL, SEQUENCE FROM N. A.)  
 RA TEDDER T. F., ISAACS C. M., ERNST T. J., DEMETRI G. D., ADLER D. A.,  
 RA DISTECHE C. M.;  
 RL J. EXP. MED. 170:123-133(1989).  
 DR EMBL; X16150; HSLYAM1.  
 DR PROSITE; PS00022; EGF.  
 KW CELL ADHESION; GLYCOPROTEIN; SIGNAL.  
 FT SIGNAL 1 28  
 FT PROPEP 29 38  
 FT CHAIN 39 372 LEUKOCYTE ADHESION MOLECULE-1.  
 FT CARBOHYD 104 104 POTENTIAL.  
 FT CARBOHYD 177 177 POTENTIAL.  
 FT CARBOHYD 232 232 POTENTIAL.  
 FT CARBOHYD 246 246 POTENTIAL.  
 FT CARBOHYD 271 271 POTENTIAL.  
 SQ SEQUENCE 372 AA; 42313 MW; 724484 CN;

Initial Score = 9 Optimized Score = 49 Significance = 4.12  
 Residue Identity = 22% Matches = 66 Mismatches = 174  
 Gaps = 50 Conservative Substitutions = 0

```

      X      10      20      30      40      50
      MGNNCYNVVVIVLLL VGCEKVG---AVQNSCDNCQPGTFCRKYNPVCKSCPPSTFSSIGGQP
      :      :      :      :      :      :      :      :      :
AEIEYLEKTLPF SRSYYWIGIRKIGGIWTWVG TNKSLTEE AENWGDGEPNNKKNKEDCVEIYIKRNKDAGKW
      80      90      100      110      120      130      140

60      70      80      90      100      110
N---CNICRV CAGYFRFKKFCSS THNAEC-ECIEGFHC-----LGPQC---TRCEKDCRPGQELTKQGCKT
:      :      :      :      :      :      :      :      :      :
NDDACHKLKAALCYTASCQPWSCSGHGECVEI INNYTCNCDVGYYPQCQFVIQCEPLEAP--ELGTMDCT
      150      160      170      180      190      200      210

120      130      140      150      160      170      180
CSLGTFN-DQNGTGVC RPWTNCSLDGRSVLKTGTTEKDVVCGPPVVSFSPSTTISV-----TPEGGPGGH
:      :      :      :      :      :      :      :      :      :
HPLGNFNFNSQCAFSCSEGTN--LTG----IEETT-----CEPFGNWSSPEPTCQVIQCEPLSAPDLGIMNC
      220      230      240      250      260      270

      190      200      210      220      230      240
SLQVLTLFLALTSALLLALIFITLLFSVLKWI-----RKKFPHIFKQPFKKTGAAQEEEDACSCRCP-----
:      :      :      :      :      :      :      :      :      :
S-HPLASF-SFTSACTFICSEGTELIGKKKTICESSGIWSNPSPICQKLDKSFSMIKEGDYNPLFIPVAVMV
      280      290      300      310      320      330      340

      250      X
QEEEGGGGGYEL
      :      :
TAFSGLAFIIWLARRLKKGKKS
      350      X      360

```



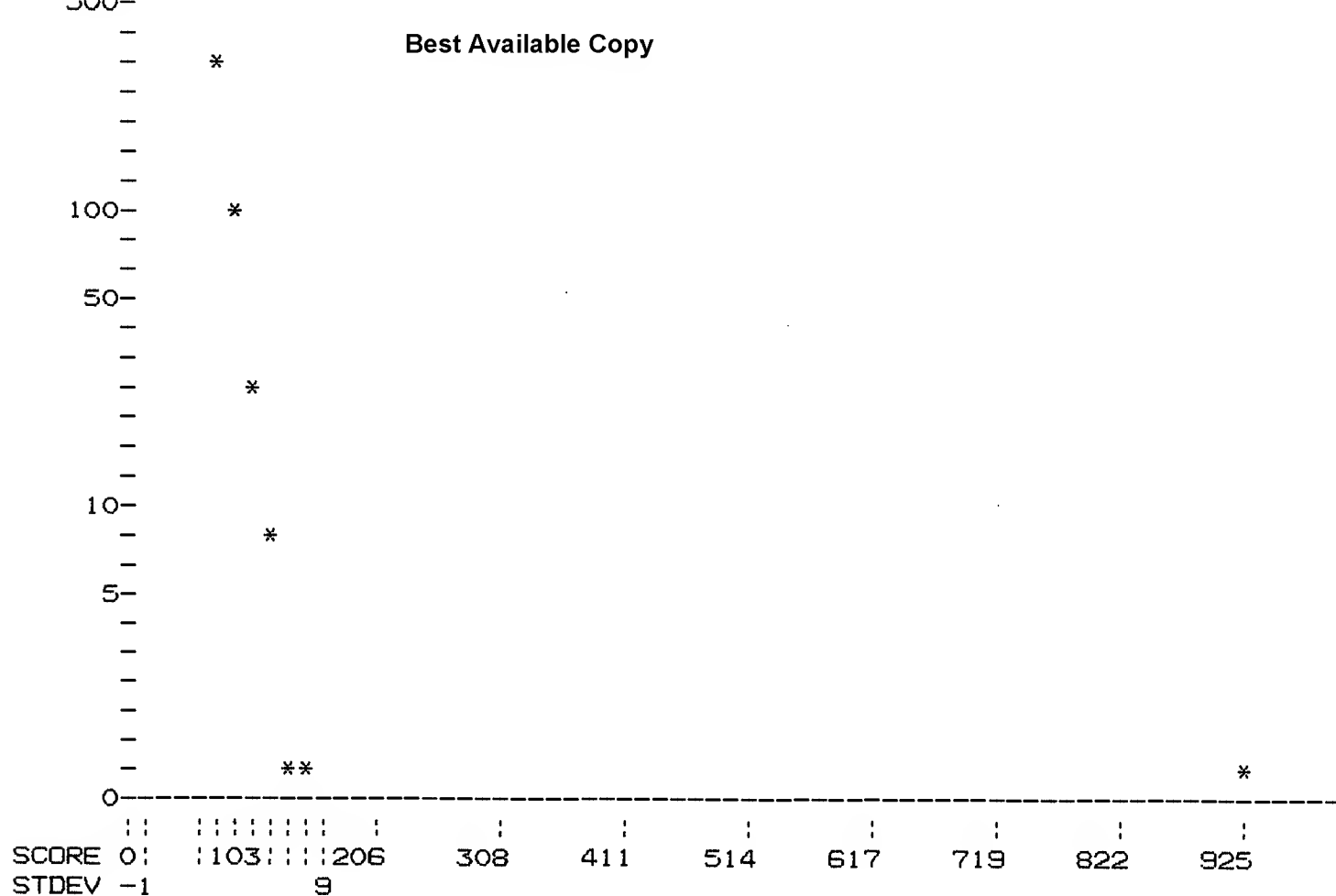
Results file ellis-267-3a.res made by wendyc on Mon 27 Aug 90 14:56:51-PDT.

Query sequence being compared: ELLIS-267-3A  
Number of sequences searched: 39513  
Number of scores above cutoff: 2415

Results of the initial comparison of ELLIS-267-3A with:  
Data bank : GenBank 64.0, all entries  
Data bank : UEMBL 23\_64.0, all entries

100000-  
-  
N -  
U50000-  
M -  
B -  
E -  
R - \*  
- \*  
O -  
F10000-  
-  
S - \*  
E 5000-  
Q -  
U -  
E -  
N \*  
C - \*  
E -  
S 1000-  
-  
-

# Best Available Copy



## PARAMETERS

Similarity matrix	Unitary	K-tuple	4
Mismatch penalty	1	Joining penalty	30
Gap penalty	1.00	Window size	32
Gap size penalty	0.33		
Cutoff score	55		
Randomization group	0		
Initial scores to save	20	Alignments to save	10
Optimized scores to save	20	Display context	10

## SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	33	33	13.33
Times:	CPU	Total Elapsed	
	00:41:02.99	00:56:42.00	
Number of residues:	49483801		
Number of sequences searched:	39513		
Number of scores above cutoff:	2415		

The scores below are sorted by initial score.  
Significance is calculated based on initial score.

A 100% similar sequence to the query sequence was found:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
---------------	-------------	--------	-------------	------------	------	-------

1. MUSTC41BB Mouse T-cell receptor 4-1BB pr 2350 925 925 66.92 0  
Best Available Copy

The list of other best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
**** 8 standard deviations above mean ****						
2. HUMCS3	Human chorionic somatomammotro	2740	147	401	8.55	0
**** 7 standard deviations above mean ****						
3. NEULCC	N. crassa laccase gene, 3'end.	726	135	310	7.65	0
**** 6 standard deviations above mean ****						
4. MNKHBD	Spider monkey (A. geoffroyi) de	1959	123	372	6.75	0
5. TOGTBESP	Tick-borne-encephalitis virus	2450	120	398	6.53	0
6. SEHCRYAA1	Mole rat alpha-A-crystallin ge	5491	119	329	6.45	0
7. HUMGHCSA	Human growth hormone (GH-1 and	66495	118	401	6.38	0
8. HUMNRASR	Human N-ras mRNA and flanking	2436	117	395	6.30	0
9. RSNEU	Rat mRNA for neuraxin	3418	116	405	6.23	0
10. RATFAS	Rat mRNA for fatty acid syntha	8936	116	415	6.23	0
11. RATFAST	Rat fatty acid synthetase mRNA	2805	115	414	6.15	0
12. DRETUBB2	D. melanogaster beta-2 tubulin	1403	113	396	6.00	0
**** 5 standard deviations above mean ****						
13. MUSAB321	Mouse MHC A-beta-3/A-beta-2 me	2689	112	396	5.93	0
14. PIGUFMR	Pig uteroferrin mRNA, complete	1424	110	403	5.78	0
15. M22618	Figure 3. Nucleotide sequence	7253	108	396	5.63	0
16. HSHGMCSF	Human mRNA for granulocyte-mac	1807	108	398	5.63	0
17. HUMCYPMP	Human liver cytochrome P-450 S	1576	108	353	5.63	0
18. M27685	Figure 2. The nucleotide seque	1717	108	404	5.63	0
19. MZEZE19B1	Maize 19 kDa zein mRNA, clone	852	108	288	5.63	0
20. HUMCYPMPA	Human cytochrome P-450 S-mephe	1577	108	353	5.63	0

The scores below are sorted by optimized score.

Significance is calculated based on optimized score.

A 100% similar sequence to the query sequence was found:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
1. MUSTC41BB	Mouse T-cell receptor 4-1BB pr	2350	925	925	91.05	0

The list of other best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
**** 5 standard deviations above mean ****						
2. MZEPOD	Maize pyruvate,orthophosphate	3171	59	424	5.14	0
**** 4 standard deviations above mean ****						
3. MUSB3RP	Mouse band 3-related protein m	4088	59	422	4.80	0
4. RATTPOFR	Rat thyroid peroxidase (TPO)	2777	62	420	4.46	0
5. RNTPO	Rat mRNA for thyroid peroxidas	3237	62	420	4.46	0
6. FLAP1M	Influenza A/nt/60/68 (h3n2), p	2341	57	419	4.29	0
7. ECOORI	E. coli replication origin (ori	2675	67	418	4.12	0
8. ECOORIASN	E. coli replication origin (ori	4012	67	418	4.12	0
9. RATBAND33E	Rat band 3 C1-/HCO3- exchanger	4057	55	418	4.12	0
10. ECASNA	E. coli asn-A gene for asparag	2170	67	418	4.12	0
11. HUMHBA4	Human alpha globin psi-alpha-1	12847	63	418	4.12	0
12. MUSADAM	Mouse adenosine deaminase mRNA	1379	64	418	4.12	0
**** 3 standard deviations above mean ****						
13. HUMINSR	Human insulin receptor mRNA, c	4723	72	417	3.94	0
14. CHKERBBF	Chicken c-erbB oncogene mRNA a	6563	99	417	3.94	0
15. HAMAPRTG	Hamster apt gene for adenine	3960	57	417	3.94	0
16. HUMNCAM	Human neural cell adhesion mol	1423	72	417	3.94	0

17. HUMINSRA	Human insulin receptor mRNA, c	5173	72	417	3.94	0
18. HUMPDGFRA	Human platelet-derived growth	5570	60	416	3.77	0
19. FLAPBIAC	Influenza A/Mallard/New York/6	2341	56	416	3.77	0
20. HUMALDC	Human aldolase C gene.	4252	63	416	3.77	0

# 1. ELLIS-267-3A

MUSTC41BB Mouse T-cell receptor 4-1BB protein mRNA, complete

LOCUS MUSTC41BB 2350 bp ss-mRNA ROD 15-SEP-1989  
 DEFINITION Mouse T-cell receptor 4-1BB protein mRNA, complete cds.  
 ACCESSION J04492  
 KEYWORDS T-cell receptor.  
 SOURCE Mouse (strain C57BL/6) T-lymphocyte cell lines L2 and L3, cDNA to mRNA.  
 ORGANISM Mus musculus  
 Eukaryota; Animalia; Metazoa; Chordata; Vertebrata; Mammalia; Theria; Eutheria; Rodentia; Myomorpha; Muridae; Murinae; Mus; musculus.  
 REFERENCE 1 (bases 1 to 2350)  
 AUTHORS Kwon,B. S. and Weissman,S. M.  
 TITLE cDNA sequences of two inducible T-cell genes  
 JOURNAL Proc. Natl. Acad. Sci. U. S. A. 86, 1963-1967 (1989)  
 STANDARD full staff\_review  
 COMMENT Draft entry and clean copy of sequence for [1] kindly provided by B. S. Kwon, 17-MAR-1989.  
 FEATURES Location/Qualifiers  
 CDS 146..916  
 /note="4-1BB protein precursor"  
 sig\_peptide 146..214  
 /note="4-1BB protein signal peptide"  
 mat\_peptide 215..913  
 /note="4-1BB protein"  
 BASE COUNT 590 a 561 c 589 g 607 t 3 others  
 ORIGIN Unreported.

Initial Score = 925 Optimized Score = 925 Significance = 91.05  
 Residue Identity = 100% Matches = 925 Mismatches = 0  
 Gaps = 0 Conservative Substitutions = 0

```

X      10      20      30      40      50      60      70
ATGTCCATGAACTGCTGAGTGGATAAACAGCACGGGATATCTCTGTCTAAAGGAATATTACTACACCAGGAA
|||||
ATGTCCATGAACTGCTGAGTGGATAAACAGCACGGGATATCTCTGTCTAAAGGAATATTACTACACCAGGAA
X      10      20      30      40      50      60      70

      80      90     100     110     120     130     140
AAGGACACATTTCGACAACAGGAAAGGAGCCTGTACAGAAAACACAGTGTCTGTGCATGTGACATTTTCGC
|||||
AAGGACACATTTCGACAACAGGAAAGGAGCCTGTACAGAAAACACAGTGTCTGTGCATGTGACATTTTCGC
      80      90     100     110     120     130     140

     150     160     170     180     190     200     210
CATGGGAAACAACACTGTTACAACGTGGTGGTCATTGTGCTGCTAGTGGGCTGTGAGAAGGTGGGAGCCGT
|||||
CATGGGAAACAACACTGTTACAACGTGGTGGTCATTGTGCTGCTAGTGGGCTGTGAGAAGGTGGGAGCCGT
     150     160     170     180     190     200     210

     220     230     240     250     260     270     280
GCAGAACTCCTGTGATAACTGTCAGCCTGGTACTTTCTGCAGAAAATACAATCCAGTCTGCAAGAGCTGCCC
|||||
GCAGAACTCCTGTGATAACTGTCAGCCTGGTACTTTCTGCAGAAAATACAATCCAGTCTGCAAGAGCTGCCC
     220     230     240     250     260     270     280

     290     300     310     320     330     340     350     360
TCCAAGTACCTTCTCCAGCATAGGTGGACAGCCGAAGTGTAAACATCTGCAGAGTGTGTGCAGGCTATTTTCAG

```

Best Available Copy

```

TCCAAGTACCTTCTCCAGCATAGGTGGACAGCCGAACCTGTAACATCTGCAGAGTGTGTGCAGGCTATTTTCAG
290      300      310      320      330      340      350      360

      370      380      390      400      410      420      430
GTTCAAGAAGTTTTGCTCCTCTACCCACAACGCGGAGTGTGAGTGCATTGAAGGATTCCATTGCTTGGGGCC
.....
GTTCAAGAAGTTTTGCTCCTCTACCCACAACGCGGAGTGTGAGTGCATTGAAGGATTCCATTGCTTGGGGCC
      370      380      390      400      410      420      430

      440      450      460      470      480      490      500
ACAGTGCACCAGATGTGAAAAGGACTGCAGGCCTGGCCAGGAGCTAACGAAGCAGGGTTGCAAAACCTGTAG
.....
ACAGTGCACCAGATGTGAAAAGGACTGCAGGCCTGGCCAGGAGCTAACGAAGCAGGGTTGCAAAACCTGTAG
      440      450      460      470      480      490      500

      510      520      530      540      550      560      570
CTTGGGAACATTTAATGACCAGAACGGTACTGGCGTCTGTGACCCCTGGACGAACTGCTCTCTAGACGGAAG
.....
CTTGGGAACATTTAATGACCAGAACGGTACTGGCGTCTGTGACCCCTGGACGAACTGCTCTCTAGACGGAAG
      510      520      530      540      550      560      570

      580      590      600      610      620      630      640
GTCTGTGCTTAAGACCGGGACCACGGAGAAGGACGTGGTGTGTGGACCCCTGTGGTGAGCTTCTCTCCAG
.....
GTCTGTGCTTAAGACCGGGACCACGGAGAAGGACGTGGTGTGTGGACCCCTGTGGTGAGCTTCTCTCCAG
      580      590      600      610      620      630      640

      650      660      670      680      690      700      710      720
TACCACCATTTCTGTGACTCCAGAGGGAGGACCAGGAGGGCACTCCTTGACGGTCCTTACCTTGTTCTGGC
.....
TACCACCATTTCTGTGACTCCAGAGGGAGGACCAGGAGGGCACTCCTTGACGGTCCTTACCTTGTTCTGGC
      650      660      670      680      690      700      710      720

      730      740      750      760      770      780      790
GCTGACATCGGCTTTGCTGCTGGCCCTGATCTTCATTACTCTCCTGTTCTCTGTGCTCAAATGGATCAGGAA
.....
GCTGACATCGGCTTTGCTGCTGGCCCTGATCTTCATTACTCTCCTGTTCTCTGTGCTCAAATGGATCAGGAA
      730      740      750      760      770      780      790

      800      810      820      830      840      850      860
AAAATTCCCCACATATTCAAGCAACCATTTAAGAAGACCACTGGAGCAGCTCAAGAGGAAGATGCTTGTAG
.....
AAAATTCCCCACATATTCAAGCAACCATTTAAGAAGACCACTGGAGCAGCTCAAGAGGAAGATGCTTGTAG
      800      810      820      830      840      850      860

      870      880      890      900      910      920      X
CTGCCGATGTCCACAGGAAGAAGAAGGAGGAGGAGGAGGCTATGAGCTGTGATGTACTATC
.....
CTGCCGATGTCCACAGGAAGAAGAAGGAGGAGGAGGAGGCTATGAGCTGTGATGTACTATCCTAGGAGATG
      870      880      890      900      910      920      X      930

```

## 2. ELLIS-267-3A

MZEPOD Maize pyruvate,orthophosphate dikinase mRNA, compl

LOCUS MZEPOD 3171 bp ss-mRNA PLN 30-SEP-1988  
 DEFINITION Maize pyruvate,orthophosphate dikinase mRNA, complete cds.  
 ACCESSION J03901  
 KEYWORDS pyruvate,orthophosphate dikinase.  
 SOURCE Maize (strain Golden Cross Bantam) green leaf, cDNA to mRNA, clones pPPD[71,1067].  
 ORGANISM Zea mays  
 Eukaryota; Plantae; Embryobionta; Magnoliophyta; Liliopsida;  
 Commelinidae; Cyperales; Poaceae; Zea; mays.  
 REFERENCE 1 (bases 1 to 3171)





CDS

183..3896  
Best Available Copy

/note="band 3-related protein"

repeat\_region 3882..3938

/note="degenerate tandem repeat copy A"

repeat\_region 3939..4001

/note="degenerate tandem repeat copy B"

BASE COUNT 842 a 1186 c 1211 g 849 t

ORIGIN 106 bp upstream of XbaI site.

Initial Score = 59 Optimized Score = 422 Significance = 4.80  
Residue Identity = 50% Matches = 504 Mismatches = 368  
Gaps = 124 Conservative Substitutions = 0

```

      X      10      20      30      40      50
      ATGTCCATGAACTGCTGAGTGGATAAACAGCACGGGATATCTCTG--TCTAAAGGA-ATATT
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
CACCTCGGGCACGGCCACGGGCCCC-----GCATAAGCCCATGAGGTGTTTGTGGAGCTGAATGAGCTGCT
1100      1110      1120      1130      1140      1150      1160

60      70      80      90      100      110      120
ACTACACCAGGAAAAGGA--CACATTCGACAACAGGAAAGGAGCCTGTCACAGAAAACCACAGTGTCTGTG
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
GTTGGACAAAACCAGGAGCCTCAGTGG----CGGGAGA-CAGCCCGCTGGATAAAATTTCGAGGAGGATGTG
1170      1180      1190      1200      1210      1220

130      140      150      160      170      180      190
CATG-TGACATTTGCCATGGGAAACAACCTGTTACAACGTGGTGGTCATTGTGCTG-CTGCTAGTGGGCTGT
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
GAAGAGGAGACTGAGCGCTGGGGGAAGCCTCATGTGGCCTCACTGTCCTTCCGTAGCCTCCTGGAGCTCCGC
1230      1240      1250      1260      1270      1280      1290      1300

200      210      220      230      240      250      260
GAGAAGGTGGGAGCCGTGCAGAACTCCTGTGATAACT--GTCAGCCTGGTACTTTCTGCAGAAAATACAATC
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
AGGACTCTGG--CCCATGGAGCTGTGCTCTTAGACCTCGATCAG-CAGACCCTGCCTG--GGGTGGCCCATC
1310      1320      1330      1340      1350      1360

270      280      290      300      310      320      330
CAGTCTGCAAG-AGCTGCCCTCCAAGTACCTTCTCCA-GCATAGGTGGACAGCCGAACCTGTAAC---ATCTG
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
AGGTGGTTCGAGCAGATGGTCATCTCTGACCAGATCAAGGCAGAGG-ATAGAGCCAATGTGCTACGGGCCCTC
1370      1380      1390      1400      1410      1420      1430

340      350      360      370      380      390      400
CAGAGTGTGTGCAGGCTATTTCAGGTTCAAGAAGTTTTGCTCCTCTACCCACAAC-----GCGGAGTGT-G
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
CTGCTAAAGCACA-GCCACCCAAGTGACGAGAAAGAGTTTCTCCTTCCCCGAAACATCTCAGCGGGCTCTCT
1440      1450      1460      1470      1480      1490      1500      1510

410      420      430      440      450      460
AG----TGCAATTGAAGGAT-TCCATTGCTTGGGG--CCACAGTGCACCAGATGTGAAAAGGACT-GCAGGCC
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
AGGCTCTCTACTGGGGCATCACCATGCCAGGGGACCGAGAGTGATCCTCATGTCACTGAGCCTCTCATTGG
1520      1530      1540      1550      1560      1570      1580

470      480      490      500      510      520      530
TGG----CCAG-GAGCTAAC-GAAGCAGGGTTGCAAAACCTGTAGCTTGGGAACATTTAATGACCAGAACGG
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
TGGTGTTCCTGAGACCCGACTGGAGGTGGATAG-AGAGCGTG-AGCTACCACCCCCAGCACCACCTGCA-GG
1590      1600      1610      1620      1630      1640      1650

540      550      560      570      580      590      600
TACTGGCGTCT-GTCGACCTGGACGAACTG-CTCT-CTAGACGGAAGGTCTGTGCTTAAGACCGGGACCAC
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
TATTACCCGCTCCAAGTCCAAGCATGAGCTGAAGCTGCTGGA--GAAGATCCCTG-AGAATGCGGAGGCTAC
```



Best Available Copy

```

      610      620      630      640      650      660
GGAGAAGGACGTGGTGTGTGGACCCCTGTGGTGAGCTTCTCTCCAGTACCACCAT-----TTCTGTGACTC
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
AG---TGGTCCTCGTGGG-----CTGTGTGGAGTTCCTCTCCC-GCCCTACCATGGCCTTC-GTG-CGG
      1730      1740      1750      1760      1770

670      680      690      700      710      720      730
CAGAGGGAGGACCAGGAG-GGCACTCCTTGCAAGGTCCTTACCTTGTTCTTG-GCGCTGACATCGGCTTTGCT
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
TTGCGGGAGGCTGTGGAGCTGGATGCCGTGCTAG-AGGTGCCT-GTGCCTGTGCGCT-TCCTCTTC-TTGCT
1780      1790      1800      1810      1820      1830      1840

740      750      760      770      780      790      800
GCTGG--CCCTGATCTTCATTACTCTCCTGTTCT-CTGTGCTCAAATGGATCAGGAAAAAATTCC-CCCACA
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
GCTGGGACCCAG-----CAGTGCTAACATGGACTACCATG---AGATCGGCC--GCTCCATTTCCACCCTCA
1850      1860      1870      1880      1890      1900

      810      820      830      840      850      860      870
TAT---TCAAGCAA--CCATTTAAGAAGACCA-CTGGAGCA-GCTCAAGAGGAAGATGCTTGTAGCTGCCGA
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
TGTCTGACAAGCAATTTCA-TGAGGCAGCCTACCTGGCGGATGAACGAGACG-ACTTGCTGACTGCTATCAA
1910      1920      1930      1940      1950      1960      1970

      880      890      900      910      920      X
TGTC-CACAGGA-AGAAG-----AAGG---AGGAGGAGGA-GGCTATGAGCTGTGATGTACTATC
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
TGCCTTCCTGGACTGCAGTGTGTGCTACCGCCTTCTGAAGTGCAGGGCGAGGAGCTGCTGCGTTCTGTTGC
1980      1990      2000      2010      2020      2030      2040      2050

CCATTTCC

```

#### 4. ELLIS-267-3A

RATTPOFR Rat thyroid peroxidase (TPO) mRNA, 3' end.

LOCUS	RATTPOFR	2777 bp ss-mRNA	ROD	15-JUN-1990
DEFINITION	Rat thyroid peroxidase (TPO) mRNA, 3' end.			
ACCESSION	M31655			
KEYWORDS	thyroid peroxidase.			
SOURCE	Rat thyroid cell line FRTL-5, cDNA to mRNA.			
ORGANISM	Rattus norvegicus			
	Eukaryota; Animalia; Metazoa; Chordata; Vertebrata; Mammalia;			
	Theria; Eutheria; Rodentia; Myomorpha; Muridae; Murinae; Rattus;			
	norvegicus.			
REFERENCE	1 (bases 1 to 2777)			
AUTHORS	Isozaki, O., Kohn, L. D., Kozak, C. A. and Kimura, S.			
TITLE	Thyroid peroxidase: Rat cDNA sequence, chromosomal localization in mouse, and regulation of gene expression by comparison to thyroglobulin in rat FRTL-5 cells			
JOURNAL	Mol. Endocrinol. 3, 1681-1692 (1989)			
STANDARD	simple staff_entry			
FEATURES	Location/Qualifiers			
CDS	(1..2313			
	/note="thyroid peroxidase"			
BASE COUNT	690 a	752 c	722 g	613 t
ORIGIN				

Initial Score	=	62	Optimized Score	=	420	Significance	=	4.46
Residue Identity	=	51%	Matches	=	515	Mismatches	=	347
Gaps	=	137	Conservative Substitutions	=			=	0

X	10	20	30	40	50
---	----	----	----	----	----

ACAGACGCTCAGAGGCAGGACTGGAHAGCATTCACTACCTCGGGTCATCTGTGACAACACCGGCCTCAC  
1570 1580 1590 1600 1610 1620 1630

60 70 80 90 100 110  
TACTACACCAGGAAAAG---GACACATTCTGA-----CAACAGGA--AAGGAGCCTGTCACAGAAAACCACA  
CAGAGTACCTGTGGATGCCTTCCGTATTGGAAAGTTCCCCAGGACTTTGAATCCTGTGA--GGAAATCCCTA  
1640 1650 1660 1670 1680 1690 1700

120 130 140 150 160 170 180  
GTGTCCTGTGTCATGTGACATTTGCCATGGGAAACAACCTGTTACAACGTGGTGGTCATTGTG-CTGC----T  
G---CATGGACCTCAGAC-TGTGG---AGGGAGAC--CT--TCCCACAAGACGACAAGTGTGTCTTCCCAGA  
1710 1720 1730 1740 1750 1760

190 200 210 220 230 240 250  
GCTAGTGGGCTGTGAGAA--GGTGGGAGCCGTGCAGAACTCCTGTGATAACTGTCAGCCTGGTACTTTCTGCA  
GAAGGTGGACAATGGGAACTTTGTGCACTGTGAAGAA--TC-TGGGA-AGCTGGTA--CTGGTGTATTCTGT  
1770 1780 1790 1800 1810 1820 1830

260 270 280 290 300 310  
GAAAATACAATCCAGTCTGCAAGAGCTGCCCTCCAAGT-ACCT-TCTCCAGCATAGGT-GGAC--AG-----  
TTCCAT-GGATACAAGCTGCAAGGCCAG--GAGCAGGTACATGTACCCAGAATGGATGGGACTCAGAGCCT  
1840 1850 1860 1870 1880 1890 1900

320 330 340 350 360 370 380 390  
CCGAAGTGTAACTCTG-CAGAGTGTGTGCAGGCTATTTCAAGTTCAGAAGTTTTGCTCCTCTACCCACAA  
CCTGTCTGTAA-AGATGTTAATGAGTGTGCAG---AT--CTGACACACCCACCT-TGCCACTC-CTCCGCAA  
1910 1920 1930 1940 1950 1960

400 410 420 430 440 450  
CGCGGAGTGTGAGTGCATTGAAGGA---TTCCATTGCTTGGGGCCACAGTGCACCAGATG--TGAAAAGGA-  
-----AGTGCAAGAACACCAAGGGAAGCTTCCAGTG--TGTGTGCACAGACCCCTACATGCTAGGTGAGGAT  
1970 1980 1990 2000 2010 2020 2030

460 470 480 490 500 510 520  
CTGCAGGCCTGGCCAGGAGCTAACGAAGCAGGGTTGCAAAACCTGTAGCTTGGGAACATTTAATG-AC--CA  
GAGAAGACCTGCATAGATTCTGGC-AGGCTACCTCGGGCATCCTGGGTCTCCATTGCATTGGGTGCACTTCT  
2040 2050 2060 2070 2080 2090 2100

530 540 550 560 570 580 590  
GAACGGTACTGGCGTCTGTGACCCTGGACGAACCTGCTCTCTAGACGGA-AGGT-CTGTGCTTAAGACC---  
CATTGGTGGTTTGGCCAGTCTCAGCTGGAC--TGTAATTTGCAGGTGGACACATGCTGATAAGAAGTCCACA  
2110 2120 2130 2140 2150 2160 2170

600 610 620 630 640 650  
---GGGACCACGGAGA-AG-GA-CGTGGTGTGTGGACCCCCTGTG-GTGAGCTTCTCTCCAGT--ACCAC-  
TTGCTGATCACCGAGAGAGTGACCATGGAGTCAGGATTCAGAAAGAGTCAG-GAGAGTGGGATTTACCCACA  
2180 2190 2200 2210 2220 2230 2240

660 670 680 690 700 710 720  
CATTTCTGTGACTCCAGAGGGAGGACCAGGA--GGGC-ACTCCTTGAG-GTCCTTACCT-TGTTCTTGCG  
AAAGGCCGAGGTTCAAGA-TGCTGAACAGGAACCGGCTTATGGATCCAGAGTCCT--CCTGTGTGAATAGAA  
2250 2260 2270 2280 2290 2300 2310

730 740 750 760 770 780 790

CTGACATCGGCTTTGCTGCTGGCCCTGATCTTCATT--ACTCTCCTGTCTCTGTGCTCAAAATGGATCAGGA  
 GTCCTCACTGCTTTGGAGCCAGACATTGGC-TAATTCAAGTCTCAAGCTGCCTGGG--CAAA--GA--AAGA  
 2320 2330 2340 2350 2360 2370 2380

800 810 820 830 840 850 860  
 AAAAATTCCCCACATATTCAAG-CAACCATTTAAGAAGACCA-CTGGAGCAGCTCAAGAGGAAGATGCT-T  
 CATGAT-----ACATGTTGAAGTCAGAGGCTTGAGGACACCAGATGGTTAATCTTATCAGTCCAAGGCTGC  
 2390 2400 2410 2420 2430 2440

870 880 890 900 910 920  
 GTAGCTGCCGATGTCCA---CAGGAAGAAGAAGGAGGAGG--AGGCTATGAGCTGTG---ATGTACTAT  
 ATAGCT--GAGTTCCATCTCATGTTTTTCCA-CAGGAGCAGGCCAGGCCA-GA-CTGTGCTAATG-CCTCT  
 2450 2460 2470 2480 2490 2500 2510

X  
 C  
 :  
 CCTACACAAGT  
 X 2520

## 5. ELLIS-267-3A

RNTPO Rat mRNA for thyroid peroxidase

ID RNTPO standard; RNA; ROD; 3237 BP.  
 XX  
 AC X17396; M27275;  
 XX  
 DT 05-JAN-1990 (annotation)  
 XX  
 DE Rat mRNA for thyroid peroxidase  
 XX  
 KW thyroid peroxidase.  
 XX  
 OS Rattus norvegicus (rat)  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Tetrapoda; Mammalia;  
 OC Eutheria; Rodentia.  
 XX  
 RN [1] (bases 1-3237)  
 RA Rapoport B.;  
 RT ;  
 RL Submitted (22-AUG-1989) on tape to the EMBL Data Library.  
 XX  
 RN [2] (bases 1-3237)  
 RA Derwahl M., Seto P., Rapoport B.;  
 RT "Complete nucleotide sequence of the cDNA for thyroid peroxidase  
 RT in FRTL5 rat thyroid cells";  
 RL Nucleic Acids Res. 17:8330-8330(1989).  
 XX  
 CC \*source: cell line=FRTL5.  
 XX  
 FH Key From To Description  
 FH  
 FT CDS 42 2783 thyroid peroxidase (AA 1-914)  
 XX  
 SQ Sequence 3237 BP; 816 A; 874 C; 831 G; 716 T; 0 other;

Initial Score = 62 Optimized Score = 420 Significance = 4.46  
 Residue Identity = 51% Matches = 515 Mismatches = 347  
 Gaps = 137 Conservative Substitutions = 0

X 10 20 30 40 50  
 ATGTCCATGAACTGCTGAGTGGATAAAC-AGCACGGGATATCTCTGTCTAAA--GGAAT-AT

Best Available Copy

```
ACAGACGCTCAGAGGCAGGAAGTGGAAAGCATTCACCTCGGGTCATCTGTGACAACACCGGCCTCAC
2040      2050      2060      2070      2080      2090      2100      2110

60      70      80      90      100     110
TACTACACCAGGAAAAG---GACACATTCTGA-----CAACAGGA--AAGGAGCCTGTTCACAGAAAACCACA
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
CAGAGTACCTGTGGATGCCTTCCGTATTGGAAAGTTCCCCCAGGACTTTGAATCCTGTGA-GGAAATCCCTA
      2120      2130      2140      2150      2160      2170      2180

120     130     140     150     160     170     180
GTGTCCTGTGTCATGTGACATTTGCCATGGGAAACAACCTGTTACAACGTGGTGCTCATTGTG-CTGC----T
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
G---CATGGACCTCAGAC-TGTGG---AGGGAGAC--CT--TCCCACAAGACGACAAGTGTGTCTTCCCAGA
      2190      2200      2210      2220      2230      2240

190     200     210     220     230     240     250
GCTAGTGGGCTGTGAGAA-GGTGGGAGCCGTGCAGAACTCCTGTGATAACTGTCAGCCTGGTACTTTCTGCA
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
GAAGGTGGACAATGGGAACTTTGTGCACTGTGAAGAA-TC-TGGGA-AGCTGGTA--CTGGTGTATTCTGT
      2250      2260      2270      2280      2290      2300

260     270     280     290     300     310
GAAAATACAATCCAGTCTGCAAGAGCTGCCCTCCAAGT-ACCT-TCTCCAGCATAGGT-GGAC--AG-----
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
TTCCAT-GGATACAAGCTGCAAGGCCAG--GAGCAGGTCACATGTACCCAGAATGGATGGGACTCAGAGCCT
2310      2320      2330      2340      2350      2360      2370

320     330     340     350     360     370     380     390
CCGAAGTGTAACTCTG-CAGAGTGTGTGCAGGCTATTTTAGGTTCAAGAAGTTTTGCTCCTCTACCCACAA
||   |||||   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
CCTGTCTGTAA-AGATGTTAATGAGTGTGCAG---AT--CTGACACACCCACCT-TGCCACTC-CTCCGCAA
2380      2390      2400      2410      2420      2430      2440

400     410     420     430     440     450
CGCGGAGTGTGAGTGCATTGAAGGA---TTCCATTGCTTGGGGCCACAGTGCACCAGATG--TGAAAAGGA-
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
-----AGTGCAGAACACCAAGGGGAAGCTTCCAGTG--TGTGTGCACAGACCCCTACATGCTAGGTGAGGAT
      2450      2460      2470      2480      2490      2500

460     470     480     490     500     510     520
CTGCAGGCCTGGCCAGGAGCTAACGAAGCAGGGTTGCAAAACCTGTAGCTTGGGAACATTTAATG-AC--CA
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
GAGAAGACCTGCATAGATTCTGGC-AGGCTACCTCGGGCATCCTGGGTCTCCATTGCATTGGGTGCACTTCT
2510      2520      2530      2540      2550      2560      2570

530     540     550     560     570     580     590
GAACGGTACTGGCGTCTGTGACCCCTGGACGAAGTGTCTCTAGACGGA-AGGT-CTGTGCTTAAGACC---
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
CATTGGTGGTTTGGCCAGTCTCAGCTGGAC--TGTAATTTGCAGGTGGACACATGCTGATAAGAAGTCCACA
2580      2590      2600      2610      2620      2630      2640

600     610     620     630     640     650
---GGGACCACGGAGA-AG-GA-CGTGGTGTGTGGACCCCTGTG-GTGAGCTTCTCTCCAGT--ACCAC-
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
TTGCTGATCACCGAGAGAGTGACCATGGAGTCAGGATTCAGAAAGAGTCAG-GAGAGTGGGATTTTACCACA
2650      2660      2670      2680      2690      2700      2710

660     670     680     690     700     710     720
CATTTCTGTGACTCCAGAGGGAGGACCAGGA--GGGC-ACTCCTTGAG-GTCCTTACCT-TGTTCTTGCG
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
AAAGGCCGAGGTTCAAGA-TGCTGAACAGGAACCGGCTTATGGATCCAGAGTCCT--CCTGTGTGAATAGAA
2720      2730      2740      2750      2760      2770      2780

730     740     750     760     770     780     790
CTGACATCGGCTTTGCTGCTGGCCCTGATCTTCATT--ACTCTCCTGTTCTCTGTGCTCAAATGGATCAGGA
```

2790 2800 2810 2820 2830 2840 2850  
GTCCTCACTGCTTTGGAGCCAGACATTGGC-TAATTCAAGTCTCAAGCTGCCTGGG--CAA--GA--AGA

800 810 820 830 840 850 860  
AAAAATTCCCCACATATTCAAG-CAACCATTTAAGAAGACCA-CTGGAGCAGCTCAAGAGGAAGATGCT-T  
| | | | | | | | | | | | | | | | | | | | |  
CATGAT-----ACATGTTGAAGTCAGAGGCTTGAGGACACCAGATGGTTAATCTTATCAGTCCAAGGCTGC  
2860 2870 2880 2890 2900 2910

870 880 890 900 910 920  
GTAGCTGCCGATGTCCA---CAGGAAGAAGAAGGAGGAGGAGG--AGGCTATGAGCTGTG---ATGTACTAT  
| | | | | | | | | | | | | | | | | | | | |  
ATAGCT--GAGTTCCATCTCATGTTTTTCCA-CAGGAGCAGGCCAGGCCA-GA-CTGTGCTAATG-CCTCT  
2920 2930 2940 2950 2960 2970 2980

X  
C  
:  
CCTACACAGTA  
X 2990

6. ELLIS-267-3A

Influenza A/nt/60/68 (h3n2), polymerase 1 (seq 2),

```

LOCUS       FLAP1M             2341 bp ss-RNA             VRL             30-JUN-1987
DEFINITION  Influenza A/nt/60/68 (h3n2), polymerase 1 (seg 2), cDNA.
ACCESSION   J02138
KEYWORDS    RNA polymerase; polymerase.
SOURCE      influenza from human.
  ORGANISM  Influenza virus type A
            Viridae; ss-RNA enveloped viruses; Negative strand RNA viruses;
            Orthomyxoviridae; Influenzavirus; Influenza A viruses; Influenza
            virus type A.
REFERENCE   1 (bases 1 to 2341)
  AUTHORS   Bishop,D. H. L. , Huddleston,J. A. and Brownlee,G. G.
  TITLE     the complete sequence of rna segment 2 of influenza a/nt/60/68 and
            its encoded p1 protein
  JOURNAL   Nucleic Acids Res. 10, 1335-1343 (1982)
  STANDARD  full staff_review
COMMENT     Sequence derived from cloned cDNA (a/nt/60/2/68/1962); bases
            518-1693 also obtained independently with separate cloned cDNA
            (371). First 12 and last 13 bases questionable. Assignment of
            coding region by consideration of open reading frames.
FEATURES             Location/Qualifiers
     CDS             25..2298
                     /note="polymerase 1"
     unsure          1644
                     /note="g in clone 371; a in clone a/nt/60/2/68/1962"
BASE COUNT        827 a      460 c      530 g      524 t
ORIGIN            3' end of vRNA.

```

Initial Score	=	57	Optimized Score	=	419	Significance	=	4.29
Residue Identity	=	51%	Matches	=	516	Mismatches	=	349
Gaps	=	135	Conservative Substitutions	=	0			

X            10                  20                  30                         40                         50  
ATGTCCATGAACTGCTGAGTGGATAAACACGCACG-----GGATATCTCTGTC-TAAAGGAA  
| | | | | | | | | | | | | | | | | | | | |  
GAACACATCAATATT CAGAAAAAGGGAAGTGGACAACAAACACGGAAACTGGAGCGCCCCA ACTTAACCCAA  
160              170              180              190              200              210              220

             60                  70                  80                  90                  100                  110  
T--AT-TACTAC-ACC--AGGAAAAGGA-CACATT CGACA-ACAGGAAAGGAGCCTGT CACAGAAA ACCACA  
| | | | | | | | | | | | | | | | | | | | |

60 70 80 90 100 110  
T--AT-TACTAC-ACC--AGGAAAAGGA-CACATTGACA-ACAGGAAAGGAGCCTGTACAGAAAACCACA

[illegible]

```

      970      980      990      1000      1010      1020      1030
      Best Available Copy
860      870      880      890      900      910      920
TG--TAGCTGCC--GATGTCCACAGGAAGAAGAAGGAG-GAGGAGG---AGGCTATGAGCTGTGATGTACTA
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
AGCATCGCACCCATAATGTTCTCA--AACAAAATGGCGAGACTAGGGGAAAGGATACATGTTTCGAAAGTAAGA
      1040      1050      1060      1070      1080      1090      1100

X
TC
|
GCATGAAGCTCC
      1110

```

7. ELLIS-267-3A

EC00RI      E. coli replication origin (oriC) and asnA gene cod

LOCUS	ECOORI	2675 bp ds-DNA	BCT	15-SEP-1989
DEFINITION	E. coli replication origin (oriC) and asnA gene coding for asparagine synthetase A.			
ACCESSION	J01657 X02820			
KEYWORDS	asnA gene; asparagine synthetase; oriC gene; origin of replication; synthetase; unidentified reading frame.			
SOURCE	Escherichia coli K12 DNA.			
ORGANISM	Escherichia coli			
	Prokaryota; Bacteria; Gracilicutes; Scotobacteria; Facultatively anaerobic rods; Enterobacteriaceae; Escherichia; coli.			
REFERENCE	1 (bases 1 to 1105)			
AUTHORS	Sugimoto,K. , Oka,A. , Sugisaki,H. , Takanami,M. , Nishimura,A. , Yasuda,Y. and Hirota,Y.			
TITLE	Nucleotide sequence of Escherichia coli K-12 replication origin			
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 76, 575-579 (1979)			
STANDARD	full staff_review			
REFERENCE	2 (bases 57 to 575)			
AUTHORS	Meijer,M. , Beck,E. , Hansen,F.G. , Bergmans,H.E.N. , Messer,W. , Von Meyenburg,K. and Schaller,H.			
TITLE	Nucleotide sequence of the origin of replication of the Escherichia coli K-12 chromosome			
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 76, 580-584 (1979)			
STANDARD	full staff_review			
REFERENCE	3 (bases 506 to 2675)			
AUTHORS	Nakamura,M. , Yamada,M. , Hirota,Y. , Sugimoto,K. , Oka,A. and Takanami,M.			
TITLE	Nucleotide sequence of the asnA gene coding for asparagine synthetase of E.coli K-12			
JOURNAL	Nucleic Acids Res. 9, 4669-4676 (1981)			
STANDARD	full staff_review			
REFERENCE	4 (bases 6 to 527; 834 to 906)			
AUTHORS	Lothar,H. and Messer,W.			
TITLE	Promoters in the E.coli replication origin			
JOURNAL	Nature 294, 376-378 (1981)			
STANDARD	full staff_review			
REFERENCE	5 (bases 126 to 395)			
AUTHORS	Oka,A. , Sugimoto,K. , Sasaki,H. and Takanami,M.			
TITLE	An in vitro method generating base substitutions in preselected regions of plasmid DNA: Application to structural analysis of the replication origin of the Escherichia coli K-12 chromosome			
JOURNAL	Gene 19, 59-69 (1982)			
STANDARD	full staff_review			
REFERENCE	6			
AUTHORS	Matsui,M. , Oka,A. , Takanami,M. , Yasuda,S. and Hirota,Y.			
TITLE	Sites of dnaA protein-binding in the replication origin of the Escherichia coli K-12 chromosome			
JOURNAL	J. Mol. Biol. 184, 529-533 (1985)			

COMMENT [6] sites; dnaA binding sites.  
 Directly contributed by authors through Dr. Ooi of Kyoto Univ. The  
 422 bp region 106-527 contains ori (origin of replication), whose  
 probable left end is between 128 and 140, and whose probable right  
 end is 371 or 372. [5] reports many mutants that provided evidence  
 that ori contains special regions, spacer sequences, which separate  
 neighboring recognition sites.

FEATURES Location/Qualifiers  
 CDS complement(<1..17)  
 /note="putative 21k protein"  
 CDS complement(396..839)  
 /note="putative 16k protein"  
 CDS 1539..2531  
 /note="asparagine synthetase A (asnA)"  
 /gene="asnA"  
 misc\_RNA complement(<1..271)  
 /note="p oriL RNA transcription"  
 misc\_RNA 418  
 /note="p oriR RNA transcription (alt.)"  
 misc\_RNA 428  
 /note="p oriR RNA transcription (alt.)"  
 misc\_binding 182..197  
 /note="dnaA major binding site A [J. Mol. Biol. 184,  
 529-533 (1985)]"  
 misc\_binding 237..252  
 /note="dnaA minor binding site X1 [J. Mol. Biol. 184,  
 529-533 (1985)]"  
 misc\_binding 288..303  
 /note="dnaA major binding site B [J. Mol. Biol. 184,  
 529-533 (1985)]"  
 misc\_binding 323..338  
 /note="dnaA minor binding site X2 [J. Mol. Biol. 184,  
 529-533 (1985)]"  
 misc\_binding 362..377  
 /note="dnaA major binding site C [J. Mol. Biol. 184,  
 529-533 (1985)]"  
 conflict replace(105..105,"a")  
 /citation=[2]  
 conflict replace(105..105,"a")  
 /citation=[4]  
 conflict replace(543..545,"ac")  
 /citation=[2]

BASE COUNT 635 a 626 c 725 g 689 t  
 ORIGIN 1 bp upstream of BamHI site.

Initial Score = 67 Optimized Score = 418 Significance = 4.12  
 Residue Identity = 51% Matches = 508 Mismatches = 358  
 Gaps = 130 Conservative Substitutions = 0

```

      X      10      20      30      40      50
      ATGTCCATGAACTGCTGAGT--GGATAAACA-GCACGGGATATCTC-TGTCTAAAGG--AAT
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1430  TTCTTTTTTAATG--AATCAAAAG-TGAGTTAGGCTTTTTATTGAATGATTATTGCATGTGTGTCGGTTTTT
      1440      1450      1460      1470      1480      1490

      60      70      80      90      100      110      120
      ATTACT--ACACCAGGAAAAGGACAC---ATTCGACAACAGGAAAGGAGCCTGTACAGAAAACCACAGTG
      ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
      GTTGCTTAATCATAAGCAACAGGACGCAGGAGTATAAAAAATGAAAACCG-CTTACATTGCCAAACAACGTC
1500      1510      1520      1530      1540      1550      1560

      130      140      150      160      170      180      190
      TCCTGTGCAT-GTGACATTTTCGCCATGGGAAACAACACTGTTACAACG--TGGTGGTCATTGTGCTGCTGCTAG
      |  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
      AAATTAGCTTCGTGAAATCTCACTTTTCTCGTCAACTGGAAGAACGTCTGGGGCTGATCGAAGTCCAG-GCG
  
```



[illegible]

8. ELLIS-267-3A

ECOORIASN E.coli replication origin (oriC) and asnA gene cod

LOCUS ECOORIASN 4012 bp ds-DNA BCT 15-SEP-1989  
DEFINITION E.coli replication origin (oriC) and asnA gene coding for  
asparagine synthetase A.  
ACCESSION K00826  
KEYWORDS asnA gene; asparagine synthetase; minichromosome; oriC gene;  
origin of replication; synthetase.  
SOURCE Escherichia coli, clone (minichromosome) pCM959, DNA.  
ORGANISM Escherichia coli  
Prokaryota; Bacteria; Gracilicutes; Scotobacteria; Facultatively  
anaerobic rods; Enterobacteriaceae; Escherichia; coli.  
REFERENCE 1 (bases 1 to 4012)  
AUTHORS Buhk,H. -J. and Messer,W.  
TITLE The replication origin region of Escherichia coli: nucleotide  
sequence and functional units  
JOURNAL Gene 24, 265-279 (1983)  
STANDARD simple staff\_review  
COMMENT Plasmid pCM959 was obtained in vivo; it is a small circular  
minichromosome containing only E.coli chromosomal DNA. The  
circularisation point is at 1; this sequence represents the  
complete pCM959 sequence.  
FEATURES  
CDS Location/Qualifiers  
complement((1..589)  
/note="21K protein"  
CDS complement(968..1411)  
/note="16K protein"  
CDS complement(1501..1959)  
/note="17K protein"  
CDS 2111..3103  
/note="asparagine synthetase A"  
/gene="asnA"  
BASE COUNT 947 a 1025 c 1052 g 988 t  
ORIGIN 207 bp upstream of BglII site.

Initial Score = 67 Optimized Score = 418 Significance = 4.12  
Residue Identity = 51% Matches = 508 Mismatches = 358  
Gaps = 130 Conservative Substitutions = 0

```

      X      10      20      30      40      50
      ATGTCCATGAACTGCTGAGT--GGATAAACA-GCACGGGATATCTC-TGTCTAAAGG--AAT
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
TTCTTTTTTAAATG--AATCAAAAG-TGAGTTAGGCTTTTTATTGAATGATTATTGCATGTGTGTGCGGTTTTT
      2010 X      2020      2030      2040      2050      2060      2070

      60      70      80      90      100      110      120
ATTACT---ACACCAGGAAAAGGACAC---ATTCGACAACAGGAAAGGAGCCTGTACAGAAAACACAGTG
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
GTTGCTTAATCATAAGCAACAGGACGCAGGAGTATAAAAAATGAAAACCG-CTTACATTGCCAAACAACGTC
      2080      2090      2100      2110      2120      2130      2140

      130      140      150      160      170      180      190
TCCTGTGCAT-GTGACATTTTCGCCATGGGAAACAACACTGTTACAACG--TGGTGGTCATTGTGCTGCTGCTAG
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
AAATTAGCTTCGTGAAATCTCACTTTTCTCGTCAACTGGAAGAACGTCTGGGGCTGATCGAAGTCCAG-GCG
      2150      2160      2170      2180      2190      2200      2210

      200      210      220      230      240      250
TGGGCTGTGAGAAG-GTGGGAGCCGTGCAGAACTCCTGTGATAAC-TGTCAGCCTG-GTACTTTC--TGCAG
```

Best Available Copy

```
CCGATTCTTAGCCGTGTGGGGGATG--ACGC--AGGATAACTTGTGCGGGCTGTGAAAAAGCGGTGCAG
2220      2230      2240      2250      2260      2270

260      270      280      290      300      310      320
--AAAATACAATCCAGTCTGCAAGAGCTGCCC---TCCAA---GTACCTTCTCCAGCATAGGTGGACA-GCC
|||||  |||  |  |||||  ||  |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
GTAAAAGTGAAAGC--TCTGCCTGA--TGCCCAAGTTCGAAGTGGTTTCACTTCACTGGC-GAAGTGGAAACGTC
2280      2290      2300      2310      2320      2330      2340

330      340      350      360      370      380
-GAACT---GTAACA--TCTGCAGAGTGTGTGCAGGCTATTTCAAGTTCAGGAAGTTTTGCTCCTCTACCC
|||  |||  |||  |||||  |||  |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
AGACCTTAGGGCAACACGACTTCAGCGCGGGCGAAGGGCTGTACACGCACATGAA---AGC-CCT-TCGCC
2350      2360      2370      2380      2390      2400      2410

390      400      410      420      430      440      450
ACAACGCGGA--GTGTGAGTGCATTGAA---GGATTCCATTGCTTGGGGCCACAGTGCACCAG--ATGTGAA
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
CCGATGAAGACCGTCTTTCTCCGTTGCACTCGGTCTATGTTGACCAGTGGGACTGGGAACGCGTAATGGGCG
2420      2430      2440      2450      2460      2470      2480

460      470      480      490      500      510      520
AAGGACTGCAGGCCTGGCCAGGAGCTAACGAAGCAGGGTTG-CAAAACCTGTAGCTTGGGAACATTTAATGA
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
ACGGTGAGC-GTCAATTCTCGACTCTGA-AAAGCACGGTAGAGGCGATCTG-GGC--GGGA--ATTAAAGCA
2490      2500      2510      2520      2530      2540

530      540      550      560      570      580      590
CCAGAACGGTACTGGCGTCTGTGACCCCTGGAC--GAACTGCTCTCTAGACGGAAGGTCTGTGCTTAAGACC
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
ACCGAA-GCTGC-GGTTAGCGAAGAGTTTGGCCTGGCACCGTTC-CT-GCCGGA--TCAGATC---CACT
2550      2560      2570      2580      2590      2600

600      610      620      630      640      650      660
GGGACCACGGAGAAGGACGTGGTGTGTGGACCCCTGTGGTGAGCTTCTCTCCAGTACCACCATTTCTGTG
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
TCGTACACAG-CCAGGAGTTACTGT-----CTCGTTATCCGGATCTT-GATGCCA--AAGGGCGTGAGCG-G
2610      2620      2630      2640      2650      2660      2670

670      680      690      700      710      720      730
ACTCCAGAGGGAGGACCAGGAG-GGCACTCCTTGACAGGTCCTTACCTTGTTCTTG-GCGCT-GACATCGGCT
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
GCGATAGCGAAAGATCTTGGCGCGGTATTCTTGTCTGG-GATTGGCGGCAAGCTGAGCGATGGTCATCGCCA
2680      2690      2700      2710      2720      2730      2740

740      750      760      770      780      790      800
TTGC-TGCTGGC-CCTGATCTTCATTACTCTCCTGTTCTCTGTGCTCAAATGGATCAGGAAAAAATCCCCC
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
CGACGTGCGCGCACCGGATTATGATGA--CTGGAGCACCCCGT-CAGAGCTGGGCCATGCGGGTCTGAACGG
2750      2760      2770      2780      2790      2800      2810

810      820      830      840      850      860
ACATATTC---AAGCAACCATTTA--AGAAGA-CCACTGGAGC--AGCTCAAGAGGAAGATGCTTGTAGCT
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
CGATATTCTGGTGTGGAACCCGGTACTGGAAGATGCGTTTGAGCTTTCCTCCATGGG--GATCCGTGTAGAT
2820      2830      2840      2850      2860      2870      2880

870      880      890      900      910      920      X
GCCGA-----TG-----TCCACAGGAAGAAGAAGGAGGAGGAGGA-GGCTATGAGCTGTGATGTACTATC
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
GCCGACACGCTGAAGCATCAACTGG-CGCTGACCGGTGACGAAGATCGCCTGGAGCTG-GA-GTGGCATCAG
2890      2900      2910      2920      2930      2940      2950
GCGCTGCT
```

GCGCTGCT

## Best Available Copy

9. ELLIS-267-3A

RATBAND33E Rat band 3 Cl-/HCO3- exchanger (B3RP2) mRNA, compl

LOCUS RATBAND33E 4057 bp ss-mRNA ROD 15-JUN-1990  
 DEFINITION Rat band 3 Cl-/HCO3- exchanger (B3RP2) mRNA, complete cds.  
 ACCESSION J05166  
 KEYWORDS 3 Cl-/HCO3- exchanger.  
 SOURCE Rat stomach, cDNA to mRNA, clones RSAE[2-1,3-1].  
 ORGANISM Rattus norvegicus  
 Eukaryota; Animalia; Metazoa; Chordata; Vertebrata; Mammalia;  
 Theria; Eutheria; Rodentia; Myomorpha; Muridae; Murinae; Rattus;  
 norvegicus.  
 REFERENCE 1 (bases 1 to 4057)  
 AUTHORS Kudrycki, K. E., Newman, P. R. and Shull, G. E.  
 TITLE cDNA cloning and tissue distribution of mRNAs for two proteins that  
 are related to the band 3 Cl-/HCO3- exchanger  
 JOURNAL J. Biol. Chem. 265, 462-471 (1990)  
 STANDARD simple staff\_entry  
 FEATURES Location/Qualifiers  
 CDS 201..3905  
 /note="Cl-/HCO3- exchanger (B3RP2)"  
 mRNA 1..4057  
 /note="B3RP2 mRNA"  
 BASE COUNT 830 a 1179 c 1205 g 843 t  
 ORIGIN

Initial Score = 55 Optimized Score = 418 Significance = 4.12  
 Residue Identity = 50% Matches = 508 Mismatches = 362  
 Gaps = 137 Conservative Substitutions = 0

```

      X           10           20           30           40           50
      ATGTC---CATGAAC-TGCTGAGTGGATAAACAGCACGGGATATCTCTGT---CTAAAGGA
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
CTGGCCCCACACCTCGGGCAGCACCACGGGCCCCCATAAGCCTCATGAGGTGT-TC-GTAGAGCTGAATGA
1110      1120      1130      1140      1150      1160      1170

      60           70           80           90           100          110          120
      ATATTA-CTACACCAGGAAAAGGA--CACATTGACAAACAGGAAAGGAGCCTGTACAGAAAACCACAGTGT
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
      ATTGCAGTTGGACAAAACCAGGAGCCTCAGTGG---CGGGAGA-CAGCCCGGTGGATAAAATTTGAGGAG
1180      1190      1200      1210      1220      1230      1240

      130          140          150          160          170          180          190
      CCTGTGCATG-TGACATTTGCGCATGGGAAACAACCTGTTAC--AACGTGGTGGTCATTGTGCTG-CTGCTAG
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
      GACGTGGAAGAGGAGACTGAGCGCTGGG--GCAAGCCTCACGTGGCATCACTGTCCTTCCGCAGCCTCCTGG
1250      1260      1270      1280      1290      1300      1310

      200          210          220          230          240          250          260
      TGGGCTGTGAGAAGGTGGGAGCCGTGCAGAACTCCTGT--GATAACTGTCAGCCTGGTACTTTCTGCAGAAA
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
      AGCTCCGCAGGACACTGG--CCCATGGAGCTGTGCTCTTGGACCTCGATCAG-CAGACCCTGCCTG--GGGT
1320      1330      1340      1350      1360      1370      1380

      270          280          290          300          310          320
      ATACAATCCAGTCTGCAAG-AGCTGCCCTCCAAGTACCTTCTC-CAGCATAGGTGGACAGCCGAACTG-TA-
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
      GGCCCATCAGGTGGTCGAGCAGATGGTTATCTCTGACCAGATCAAAGCAGAGG-ACAGAGCCAATGTGCTAC
1390      1400      1410      1420      1430      1440      1450

      330          340          350          360          370          380          390
      -A---CATCTGCAGAGTGTGTGCAGGCTATTTCAAGTTTCAAGAGTTTGTCTCCTCTACCCACAAC---GCG
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
      GAGCCCTTCTGCTGAAACACAGCCACCCAAGTGATGAGAAAGAA-TTCTCCTTCCCCCGGAACATCTCAGCG

```

Best Available Copy

```

1460      1470      1480      1490      1500      1510      1520
  400      410      420      430      440      450
GAGTGTGAGTGC-----ATTGAAGGAT-TCCATTGCTTGGGGCC--ACAGTGCACCAGATGTGAAAAGGACT
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
GGCTCTCTGGGCTCTCTCCTGGGGCATCACCACGCCCGGGGACTGAGAGTGATCCTCACGTCACTGAGCCT
  1530      1540      1550      1560      1570      1580      1590

  460      470      480      490      500      510      520
-GCAGGCCTGG-----CCAG-GAGCTAAC-GAAGCAGGGTTGCAAAACCTGTAGCTTGGGAACATTTAATGAC
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
CTCATCGGTGGTGTTCCTGAGACCCGGCTGGAGGTGGATAG-AGAGCGTG-AGCTGCCGCCCCCAGCCCCAC
  1600      1610      1620      1630      1640      1650      1660

  530      540      550      560      570      580      590
CAGAACGGTACTGGCGTCT-GTCGACCCTGGACGAACTG-CTCT-CTAGACGGAAGGTCTGTGCTTAAGACC
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
CTGCA-GGTATTACCCGCTCCAAGTCCAAGCATGAGCTGAAGCTGCTGGA--GAAGATCCCTG-AGAATGCA
  1670      1680      1690      1700      1710      1720      1730

  600      610      620      630      640      650
GGGACCACGGAGAAGGACGTGGTGTGTGGACCCCTGTGGTGAGCTTCTCTCCAGTACCACCA-----TTT
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
GAGGCCACAG---TGGTCCTCGTGGG-----CTGTGTGGAGTTCTCTCCC-GCCCCACCATGGCCTTT
  1740      1750      1760      1770      1780      1790

  660      670      680      690      700      710      720
CTGTGACT-CCAGAGG--GAGGACCAGGAGGGCACTCCTTGAGGTCTTACCTTGTTCTCTG-GCGCTGACA
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
GTGCGCCTGCGGGAGGCTGTGGAAGTGA-TGCAGTAC-TGGAGGT-----GCCTGTGCCTGTGCGCT-TCC
  1800      1810      1820      1830      1840      1850

  730      740      750      760      770      780      790
TCGGCTTTGCTGCT--GGCCCTGATCTTCATTACTCTCCTGTTCT-CTGTGCTCAAATGGATCAGGAAAAAA
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
TCTTC-CTGCTGCTGGGGCCCAG-----CAGCGCCAACATGGACTACCATG---AGATTGGCC--GATCCAT
  1860      1870      1880      1890      1900      1910

  800      810      820      830      840      850      860
TTCC-CCCACATAT---TCAAGCAA--CCATTTAAGAAGACCACTGGAGCAGCTCAAGAGGAAGA--TGCTT
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
CTCCACCCTCATGTCTGACAAGCAATTCCA-CGAGGCAGCCTACCTG-GCAGATGAACGGGATGACTTGCTG
  1920      1930      1940      1950      1960      1970      1980

  870      880      890      900      910
GTAGCTGCCGATGTC-CACAGGA-AGAAG-----AAGG---AGGAGGAGGA-GGCTATGAGCTGTGA
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
ACTGCTATCAATGCCTTCCTGGACTGCAGTGTTGTGCTACCGCCTTCTGAAGTGCAGGGCGAGGAGCTGCTG
  1990      2000      2010      2020      2030      2040      2050      2060

  920      X
TGTA CTATC
| | | | |
CGTTCTGTTGCCCATTTCC
      2070

```

10. ELLIS-267-3A

ECASNA E. coli asn-A gene for asparagine-synthetase.

ID ECASNA standard; DNA; PRO; 2170 BP.

XX

AC V00263;

XX

DT 07-APR-1983 (minor modifications)

DT 09-JAN-1982 (first entry)

XX  
DE E. coli asn-A gene for asparagine-synthetase.  
XX Best Available Copy

KW synthetase.

OS Escherichia coli

OC Prokaryota; Bacteria; Gram-negative facultatively anaerobic rods;

OC Enterobacteriaceae.

XX  
RN [1] (bases 1-2170)

RA Nakamura M. , Yamada M. , Hirota Y. , Sugimoto K. , Oka A. ,

RA Takanami M. ;

RT "Nucleotide sequence of the asnA gene coding for asparagine

RT synthetase of E. coli K-12";

RL Nucleic Acids Res. 9:4669-4676(1981).

XX

FH Key From To Description

FH

FT CDS 1034 2023 reading frame asn-A

XX

SB Sequence 2170 BP; 497 A; 524 C; 606 G; 543 T; 0 other;

Initial Score = 67 Optimized Score = 418 Significance = 4.12  
Residue Identity = 51% Matches = 508 Mismatches = 358  
Gaps = 130 Conservative Substitutions = 0

```

      X      10      20      30      40      50
      ATGTCCATGAACTGCTGAGT--GGATAAACA-GCACGGGATATCTC-TGTCTAAAGG--AAT
      |||  || |||  || |||||  |||  |  |  |  |  |||  |  |||  |  |||  |
TTCTTTTTTAAATG--AATCAAAAG-TGAGTTAGGCTTTTTATTGAATGATTATTGCATGTGTGTCGGTTTTT
    930      X      940      950      960      970      980      990

      60      70      80      90      100      110      120
      ATTACT--ACACCAGGAAAAGGACAC---ATTCGACAACAGGAAAGGAGCCTGTACAGAAAACCACAGTG
      || ||  || |||  || |||||  |||  |  |  |  |||  |||  |||  |||  |||  |||
      GTTGCTTAATCATAAGCAACAGGACGCAGGAGTATAAAAAATGAAAACCG-CTTACATTGCCAAACAACGTC
    1000      1010      1020      1030      1040      1050      1060

      130      140      150      160      170      180      190
      TCCTGTGCAT-GTGACATTTTCGCCATGGGAAACAACCTGTTACAACG--TGGTGGTCATTGTGCTGCTGCTAG
      |  |||  |||  |||||  |||  |  |  |  |||||  |  ||||  |||  |||  |||  |||
      AAATTAGCTTCGTGAAATCTCACTTTTCTCGTCAACTGGAAGAACGTCTGGGGCTGATCGAAGTCCAG-GCG
    1070      1080      1090      1100      1110      1120      1130

      200      210      220      230      240      250
      TGGGCTGTGAGAAG-GTGGGAGCCGTGCAGAACTCCTGTGATAAC-TGTCAGCCTG-GTACTTTC--TGCAG
      |  |  |  |||  |||  |||||  |||  |||  |||||  |||||  |||||  |||||  |||||
      CCGATTCTTAGCCGTGTGGGGGATG-GC---ACGC--AGGATAACTTGTGCGGGCTGTGAAAAGCGGTGTCAG
    1140      1150      1160      1170      1180      1190      1200

      260      270      280      290      300      310      320
      --AAAATACAATCCAGTCTGCAAGAGCTGCCC---TCCAA---GTACCTTCTCCAGCATAGGTGGACA-GCC
      ||||  |||  |  |||||  |||  |||||  |||  |||  |||  |||  |||  |||||  |||  |||
      GTAAAAGTGAAAGC--TCTGCCTGA--TGCCAGTTTGAAGTGGTTTCACTTCACTGGC-GAAGTGGAAACGTC
    1210      1220      1230      1240      1250      1260

      330      340      350      360      370      380
      -GAACT---GTAACA--TCTGCAGAGTGTGTGCAGGCTATTTCAAGTTCAAGAAAGTTTGTCTCCTCTACCC
      || ||  |||  |||||  |||  |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
      AGACCTTAGGGCAACACGACTTCAGCGCGGGCGAAGGGCTGTACACGCACATGAA---AGC-CCT-TCGCC
    1270      1280      1290      1300      1310      1320      1330

      390      400      410      420      430      440      450
      ACAACGCGGA--GTGTGAGTGCATTGAA---GGATTCCATTGCTTGGGGCCACAGTGCACCAG--ATGTGAA
      |  |  |  |||  |||  |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
      CCGATGAAGACCGTCTTTCTCCGTTGCACTCGGTCTATGTTGACCAGTGGGACTGGGAACGCGTAATGGGCG
```

Best Available Copy

```

460      470      480      490      500      510      520
AAGGACTGCAGGCCTGGCCAGGAGCTAACGAAGCAGGGTTG-CAAAACCTGTAGCTTGGGAACATTTAATGA
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
ACGGTGAGC-GTCAATTCTCGACTCTGA-AAAGCACGGTAGAGGGCGATCTG-GGC--GGGA--ATTAAAGCA
1410      1420      1430      1440      1450      1460      1470

530      540      550      560      570      580      590
CCAGAACGGTACTGGCGTCTGTGACCCCTGGAC--GAAGTGTCTCTAGACGGAAGGTCTGTGCTTAAGACC
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
ACCGAA-GCTGC-GGTTAGCGAAGAGTTTGGCCTGGCACCGTTC-CT-GCCGGA---TCAGATC----CACT
1480      1490      1500      1510      1520      1530

600      610      620      630      640      650      660
GGGACCACGGAGAAGGACGTGGTGTGTGGACCCCTGTGGTGAGCTTCTCTCCAGTACCACCATTTCTGTG
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
TCGTACACAG-CCAGGAGTTACTGT-----CTCGTTATCCGGATCTT-GATGCCA--AAGGGCGTGAGCG-G
1540      1550      1560      1570      1580      1590

670      680      690      700      710      720      730
ACTCCAGAGGGGAGGACCAGGAG-GGCACTCCTTGAGGTCTTACCTTGTTCTCTG-GCGCT-GACATCGGCT
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
GCGATAGCGAAAGATCTTGCGCGGTATTCTTGTCGG-GATTGGCGGCAAGCTGAGCGATGGTCATCGCCA
1600      1610      1620      1630      1640      1650      1660

740      750      760      770      780      790      800
TTGC-TGCTGGC-CCTGATCTTCATTACTCTCCTGTTCTCTGTGCTCAAATGGATCAGGAAAAATTCCCCC
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
CGACGTGCGCGCACCGGATTATGATGA--CTGGAGCACCCCGT-CAGAGCTGGGCCATGCGGGTCTGAACGG
1670      1680      1690      1700      1710      1720      1730

810      820      830      840      850      860
ACATATTC---AAGCAACCATTTA--AGAAGA-CCACTGGAGC--AGCTCAAGAGGAAGATGCTTGTAGCT
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
CGATATTCTGGTGTGGAACCCGGTACTGGAAGATGCGTTTGAGCTTTCCTCCATGGG--GATCCGTGTAGAT
1740      1750      1760      1770      1780      1790      1800

870      880      890      900      910      920      X
GCCGA-----TG-----TCCACAGGAAGAAGAAGGAGGAGGAGGA-GGCTATGAGCTGTGATGTACTATC
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
GCCGACACGCTGAAGCATCAACTGG-CGCTGACCGGTGACGAAGATCGCCTGGAGCTG-GA-GTGGCATCAG
1810      1820      1830      1840      1850      1860      1870

GCGCTGCT
1880

```

macrophage inflammatory protein?  
Best Available Copy  
10310 INFLAMMATORY  
42521 PROTEIN?

2 2 MACROPHAGE INFLAMMATORY PROTEIN?  
(MACROPHAGE(W)INFLAMMATORY(W)PROTEIN?)

> d 12 1 2 ti

S PAT NO: 5,154,921 [IMAGE AVAILABLE] L2: 1 of 2  
ITILE: Promotion of maturation of hematopoietic progenitor cells

S PAT NO: 5,145,676 [IMAGE AVAILABLE] L2: 2 of 2  
ITILE: Method and agents for promoting wound healing

> d 12 1 2 bib ab ccls

S PAT NO: 5,154,921 [IMAGE AVAILABLE] L2: 1 of 2  
ATE ISSUED: Oct. 13, 1992  
ITILE: Promotion of maturation of hematopoietic progenitor cells  
NVENTOR: Ruth Sager, Brookline, MA  
Douglas Trask, Miami, FL  
Phong Le, Durham, NC  
SSIGNEE: Dana-Farber Cancer Institute, Inc., Boston, MA (U.S.  
corp.)  
PPL-NO: 07/552,746  
ATE FILED: Jul. 13, 1990  
RT-UNIT: 186  
RIM-EXMR: John J. Doll  
SST-EXMR: George C. Elliott  
EGAL-REP: Janis K. Fraser

S PAT NO: 5,154,921 [IMAGE AVAILABLE] L2: 1 of 2

STRACT:

method for promoting maturation of a hematopoietic precursor cell of an animal, which method includes the step of contacting the cell with a maturation-promoting amount of GRO, a polypeptide growth factor.  
S-CL-CURRENT: 424/93U, 93AA, 93V, 93W; 435/240.2, 240.21; 530/350, 351

S PAT NO: 5,145,676 [IMAGE AVAILABLE] L2: 2 of 2  
ATE ISSUED: Sep. 8, 1992  
ITILE: Method and agents for promoting wound healing  
IVENTOR: Thomas J. Fahey, III, New York, NY  
Barbara A. Sherry, New York, NY  
Anthony Cerami, Shelter Island, NY  
The Rockefeller University, New York, NY (U.S. corp.)  
PL-NO: 07/652,198  
ATE FILED: Feb. 7, 1991  
T-UNIT: 181  
IM-EXMR: Merrell C. Cashion, Jr.  
ST-EXMR: Choon P. Koh  
GAL-REP: Klauber & Jackson

PAT NO: 5,145,676 [IMAGE AVAILABLE] L2: 2 of 2

STRACT:

the present invention relates to the treatment of wound healing dysfunction by the administration of one or more wound healing modulators. The wound healing modulator may be selected from appropriate wound healing agents and binding partners, and particularly agents that enhance wound healing. The agent may comprise a cytokine, or mixture of cytokines that are also capable of binding to heparin, and inducing localized inflammation characterized by polymorphonuclear cell infiltration when administered subcutaneously. Particular agents comprise inflammatory cytokines MIP-1, MIP-1.alpha., MIP-1.beta. and MIP-2. Diagnostic and therapeutic utilities are proposed and pharmaceutical